External Human-Machine Interfaces on Autnomous Vehicles

Effect of message perspective and memory load on pedestrian crossing intentions

Master Thesis

By

Anne Reiff

External Human-Machine Interfaces on Autonomous Vehicles:

Effect of message perspective and memory load on pedestrian crossing intentions

By

Anne Reiff

in partial fulfilment of the requirements for the degree of

Master of Science
in Biomedical Engineering

at the Delft University of Technology

Student number:	4625994	
Project duration:	June 2019 – May 2020	
Thesis committee:	Dr.ir. Joost de Winter	TU Delft
	Ir. Yke Bauke Eisma	TU Delft
	Dr.ir. M.M. van Paassen	TU Delft

This thesis is confidential and cannot be made public until 15-05-2022

ABSTRACT

Perspective-taking is the ability to see a situation based on the viewpoint of others. In autonomous vehicle-pedestrian (AV-P) interaction, the perspective taken by the pedestrian could be affected by the design of an external Human-Machine Interface (eHMI). However, currently, there is little knowledge about the effect of message perspective on the crossing intentions of pedestrians when interpreting the intention of an AV. This study aims to investigate the effect of eHMI message perspective and cognitive load on participants' perspective-taking, as inferred from their crossing intentions. We designed a photo-based experiment and examined the effect of message perspective (egocentric (from the pedestrians' point of view): 'WALK', 'DON'T WALK' vs. allocentric: 'BRAKING', 'DRIVING' vs. ambiguous 'GO', 'STOP'), and cognitive load on the crossing intentions, response times and pupil diameter of the participans (N = 103). We added a memory task to increase the cognitive load during two-thirds of the trials, since crossing intentions can be demanding (the traffic scenario can be complex complex or the pedestrian is distracted) and therefore might influence perspective-taking.

The results showed that the egocentric messages were most persuasive as demonstrated by more uniform crossing intentions and faster response times compared to allocentric and ambiguous messages. When participants were put under cognitive load, a more efficient strategy was used to make a crossing decision as demonstrated by faster yet consistent crossing intentions compared to no memory task. No difference in cognitive load was measured for both message perspective and cognitive load at the moment of response, as evidenced by equal pupil size. Concerning the ambiguous messages, 'GO' encouraged crossing and the 'STOP' inhibited crossing, which points towards an egocentric perspective taken by the pedestrian. We conclude that pedestrians initially take an egocentric perspective if the eHMI message is ambiguous, though this egocentric bias can be overcome by using explicitly an egocentric or allocentric eHMI message perspective. In addition, we conclude that participants perform better (more uniform crossing decisions, faster responses) when the eHMI's message perspective is egocentric rather than allocentric.

ACKNOLWEDGEMENTS

First of all, I would like to express my deep gratitude to Dr. Ir. J. C. F. De Winter and Ir. Y. B. Eisma, my research supervisors, for their valuable and constructive suggestions during the planning and development of this research work, as well as their sharp visions and useful critiques whilst encouraging and inspiring me to improve my work. Furthermore, their willingness to give their time so generously has been very much appreciated. Also, I would like to thank Dr. D. Dodou and Ir. L. Kooijman, their help setting up the experiment and assisting during the experiment was invaluable. In addition, I would like to thank L. Kooijman' for sharing his knowledge about Matlab, which helped me analyzing the results. Furthermore, I would like to thank all participants for joining this experiment, which is the foundation of this work; without them there would be no data to process.

Also, I would like to thank my family and my friends for the brainstorming sessions about the experimental design and sharp and critical views on the subject as well as their encouragement and moral support.

CONTENTS

ACKNOLWEDGEMENTS **6**

1. INTRODUCTION **13**

1.1. PERSPECTIVE-TAKING 13
1.2. EXTERNAL HUMAN-MACHINE INTERFACES (EHMIS) 14
1.3. AIM OF THE STUDY 15

2. METHODS **17**

2.1. PARTICIPANTS 17
2.2. MATERIALS AND EQUIPMENT 17
2.3. INDEPENDENT VARIABLES 18
2.4. DEPENDENT VARIABLES 19
2.4.1. Keypress task 19
2.4.2. Memory task 19
2.5. PROCEDURE 19
2.6. PARTICIPANTS' TASK 20
2.7. DEPENDENT VARIABLES 20
2.7.1. Self-reported clarity 20
2.7.2. Response Behavior 20
2.7.3. Response time 21
2.7.4. Pupil diameter 21
2.8. STATISTICAL ANALYSIS 21

3. RESULTS **22**

3.1. DATA QUALITY ASSESSMENT 22
3.2. SELF-ASSESSED CLARITY 22
3.3. RESPONSE BEHAVIOR 23
3.4. RESPONSE TIMES 25
3.5. CORRELATION ANALYSIS 28
3.6. PUPIL DIAMETER 28
3.7. MEMORY TASK 31

4. DISCUSSION **33**

4.1. MAIN FINDINGS 33
4.2. EFFECT OF MESSAGE PERSPECTIVE 33
4.2.1 Clarity and crossing intentions 33
4.1.3. The perception of pedestrians towards ambiguous eHMIs 34
4.1.2 Cognitive processes 34
4.3. THE EFFECT OF MEMORY TASK 35

5. CONCLUSION **37**

6. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH **38**

7. REFERENCES **40**

8. APPENDICES **45**

APPENDIX 1: INFORMED CONSENT **46**

APPENDIX 2: STEPS AND INSTRUCTIONS **48**

APPENDIX 3: OVERVIEW OF EXPERIMENT; ONE OF THE 18 STIMULI USED IN THE EXPERIMENT: EHMI CONCEPT STOP, HIGH LOAD MEMORY TASK **50**

APPENDIX 4: OVERVIEW OF ALL STIMULI **51**

APPENDIX 5: QUESTIONNAIRE **52**

APPENDIX 6 UNPROCESSED DATA RESPONSE BEHAVIOR **55**

APPENDIX 7 LEARNING BEHAVIOR **56**
APPENDIX 8 NUMBER OF TIMES APPEARANCE OF STIMULI FOR EACH TRIAL **58**
APPENDIX 9 RESPONSE TIME AS A FUNCTION OF TRIAL NUMBER FOR EACH STIMULUS **59**
APPENDIX 10 PERCENTAGE PRESSING 'YES' AS A FUNCTION OF TRIAL NUMBER FOR EACH STIMULUS **60**
APPENDIX 11 MEAN DIFFERENCE IN RESPONSE TIMES FOR MEMORY LOAD **61**
APPENDIX 12 MEAN PUPIL DIAMETER AS A FUNCTION OF TIME FOR MESSAGE PERSPECTIVE AND EHMI **62**
APPPENDIX 13 OVERVIEW OF POST-HOC COMPARISONS OF MEMORY TASK FOR ALL TIME BINS **64**
APPENDIX 14: MEAN RESPONSE TIME (MS) FOR PRESSING 'YES' AND 'NO' FOR EACH STIMULUS **65**
APPENDIX 15: INDIVIDUAL RESULTS FOR RESPONSE TIME AND CROSSING INTENTIONS **66**

LIST OF TABLES

TABLE 1. POST HOC COMPARISONS CLARITY SCORE. 20

TABLE 2. OVERVIEW OF THE CLARITY SCORES. 22

TABLE 3. REPEATED ANOVA MEASURES FOR YIELDING AND NON-YIELDING VEHICLES FOR MESSAGE PERSPECTIVE, MEMORY LOAD AND INTERACTION BETWEEN MESSAGE PERSPECTIVE AND MEMORY LOAD 22

TABLE 4. POST HOC COMPARISONS AND THE MEAN (SD) PERCENTAGE PRESSING 'YES' FOR YIELDING VEHICLES AND PRESSING 'NO' FOR NON-YIELDING VEHICLES OF EACH GROUP WITHIN THE TESTED VARIABLE 22

TABLE 5.POST HOC COMPARISONS AMONG THE INTERACTION BETWEEN BETWEEN MESSAGE PERSPECTIVE AND MEMORY TASK FOR NON-YIELDING VEHICLES 23

TABLE 6. MEAN (SD) RESPONSE TIMES (MS) FOR YIELDING AND NON-YIELDING VEHICLES FOR EACH TRIAL. 24

TABLE 7. REPEATED ANOVA MEASURES RESPONSE TIME FOR MESSAGE PERSPECTIVE, MEMORY TASK AND INTERACTION MESSAGE PERSPECTIVE AND MEMORY TASK FOR BOTH YIELDING AND NON-YIELDING VEHICLES 24

TABLE 8. POST HOC COMPARISONS AND THE MEAN (SD) RESPONSE TIMES (MS) OF EACH GROUP WITHIN THE TESTED VARIABLE 25

TABLE 9. REPEATED ANOVA MEASURES PUPIL DIAMETER FOR MEMORY TASK FOR EACH TRIAL 27

TABLE 10. REPEATED ANOVA MEASURES PUPIL DIAMETER FOR THE FIRST SECOND OF THE ONSET OF THE IMAGE (T_{10} (T = 10 – 10,99)) AND ONE SECOND AFTER THE ONSET OF THE IMAGE (T_{11} T = 11-11.99) 28

TABLE 11. MEAN (SD) PUPIL DIAMETER (MM) FOR YIELDING AND NON-YIELDING VEHICLES FOR T_{10} AND T_{11}. 28

TABLE 12. PEARSON CORRELATION MATRIX AMONG SELECTED VARIABLES 29

TABLE 13. OVERVIEW OF THE DIFFERENCE IN RESPONSE TIMES FOR THE LOW AND THE HIGH LOAD MEMORY TASK COMPARED TOT HE BASELINE CONDITION 24

TABLE 14. MEAN RESPONSE TIME FOR EACH MESSAGE PRESSING 'YES' AND 'NO'. 25

LIST OF FIGURES

FIGURE 1. THE EXPERIMENTAL SETUP 15

FIGURE 2. ONE OF THE SIX VISUAL STIMULI USED IN THE EXPERIMENT: EHMI CONCEPT 'WALK' 16

FIGURE 3. THE EHMI CONCEPTS USED DURING THE EXPERIMENT 16

FIGURE 4. THE SEQUENCE OF A TRIAL. NOTE THAT THE TIME BETWEEN THE ONSET OF THE FIRST FIXATION CROSS AND THE ONSET OF THE IMAGE WAS IDENTICAL FOR EACH OF THE 18 TRIALS. 17

FIGURE 5. DISTRIBUTION OF ANSWERS TO THE STATEMENT 'I CAN CROSS' FOR EACH STIMULUS 21

FIGURE 6. RESPONSE TIME (IN MILLISECONDS) FOR EACH STIMULUS 23

FIGURE 7. LEFT: CORRELATION SELF REPORTED CLARITY VS CLARITY RATING, RIGHT: CORRELATION CLARITY RATING VS RESPONSE TIME 25

FIGURE 8. MEAN PUPIL DIAMETER FOR EACH MEMORY TASK AS A FUNCTION OF TIME 27

FIGURE 9. DISTRIBUTION OF ANSWERS TO THE STATEMENT 'I CAN CROSS' FOR EACH STIMULUS BEFORE PROCESSING THE DATA 52

FIGURE 10. LEARNING BEHAVIOR FOR ALL EHMIS FOR CLARITY SCORE (E.G., RESPONSE BEHAVIOR) AND RESPONSE TIME 53

FIGURE 11. LEARNING BEHAVIOR FOR MEMORY TASK FOR CLARITY SCORE (E.G., RESPONSE BEHAVIOR) AND RESPONSE TIME. 54

FIGURE 12. LEARNING BEHAVIOR FOR ALL EHMIS FOR CLARITY SCORE (E.G., RESPONSE BEHAVIOR) AND RESPONSE TIME. 23

FIGURE 13. AN OVERVIEW OF THE MEAN RESPONSE TIMES FOR EACH STIMULI FOR EACH TRIAL DURING THE EXPERIMENT. 25

FIGURE 14. AN OVERVIEW OF THE PERCENTAGE PRESSING 'YES' FOR EACH STIMULI FOR EACH TRIAL DURING THE EXPERIMENT. 27

FIGURE 15. THE EFFECT OF MESSAGE PERSPECTIVE ON THE MEAN PUPILLARY RESPONSES AS A FUNCTION OF TIME 52

FIGURE 16. THE EFFECT OF EACH EHMI CONCEPT ON THE PUPILLARY RESPONSES 53

FIGURE 17. MEAN RESPONSE TIMES FOR PRESSING 'YES' AND 'NO' FOR EACH TRIAL.. 54

FIGURE 18. INDIVIDUAL RESPONSE TIMES: OVERVIEW OF THE FIVE PARTICIPANTS WITH THE LONGEST AND FIVE PARTICIPANTS WITH THE SHORTEST REACTION TIME DURING THE EXPERIMENT 53

FIGURE 19. CROSSING INTENTIONS OF THE FIVE PARTICIPANTS WITH THE LONGEST AND FIVE PARTICIPANTS WITH THE SHORTEST REACTION TIME DURING THE EXPERIMENT TRIAL.. 54

.

1. INTRODUCTION

With the rise of autonomous vehicles (AVs), one of the major challenges reported from a human factor perspective is how automated vehicles should interact with other vulnerable road users, more specifically pedestrians (Habibovic et al., 2018). Misinterpretation appears to be one of the most common causes of pedestrian accidents (Habibovic & Davidsson, 2012).

AVs are expected to reduce accidents. However, the rise of AVs limits the opportunity for communication between a driver and a pedestrian. The driver might be distracted (e.g., reading a newspaper or talking on the phone) or absent and can, therefore, not communicate the intentions of the vehicle to the pedestrian anymore. Voids in communication between autonomous vehicles and pedestrians raise the question as to how a vehicle's intentions should be communicated in a way that is comprehensible to pedestrians and results in safe and comfortable road user interactions (Stanciu et al., 2018; Sucha, Dostal, & Risser, 2017).

For effective communication on the road to take place, it requires all road users to communicate their intentions in such a manner that it is received and understood by fellow road users. If an AV fails to communicate intent clearly or if the pedestrian misunderstands the communication attempt of the AV, this will lead to miscommunication, which can result in a hazardous situation (Stanciu et al., 2018). Finding the right balance between what, when, and how to communicate is one of the major challenges in autonomous vehicle-pedestrian (AV-P) communication design.

1.1. Perspective-taking

An integral part of communication is perspective-taking, which is defined as the ability to see a situation based on the viewpoint of others (Keysar, Barr, Balin, & Brauner, 2000). Understanding the perception of others and predicting behavior involves both an understanding of what another person is capable of doing and understanding their current goals (Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013). To see a situation from another point of view, knowledge about others' knowledge, beliefs, and intentions is required, also referred to as to their 'perspective'.

When physically interacting within an environment, our brain integrates spatial information into reference frames, which can either be egocentric or allocentric (Vukovic & Shtyrov (2017). Research on spatial perspective-taking has often been examined along two dimensions; what another person can see (level 1 perspective-taking) and how that other person sees a particular stimulus (level 2 perspective-taking). While both levels involve tracking another's perspective, level 1 perspective-taking happens automatically whereas level 2 perspective-taking involves shifting from an egocentric reference frame to an allocentric reference frame which is cognitively effortful (Flavell, Green, Flavell, Watson, & Campione, 1986). An extensive literature study regarding spatial perspective-taking suggests that observers use both levels to identify the intentions of another person (Creem-Regehr et al., 2013). An example of perspective-taking can be taken from the domain of communication. When describing, for example, the location of a vehicle (level 2 perspective-taking), the speaker can take an egocentric perspective (e.g., own point of view) by saying ‘the vehicle is to my right’ or take an allocentric perspective (e.g., listeners point of view) by saying ‘the vehicle is to your right’.

Research in perspective-taking indicates that people tend to rely on their own perspective, and adopting another perspective appears to be cognitively effortful, a phenomenon that is called an egocentric bias (Keysar, 2007; Todd, Cameron, & Simpson 2017). Failure of adjustment towards a non-egocentric perspective in human-human interaction may lead to miscommunications (Keysar & Barr, 2002). Egocentric bias has been widely documented by researchers (Davis, Conklin, Smith, & Luce, 1996; Keysar, 2007; Lin, Keysar, & Epley, 2010; Roxβnagel, C., 2000; Todd et al., 2017). Furthermore, research has shown that egocentric bias increases with time pressure and decreases with incentive accuracy (Epley, Keysar, Van Boven, & Gilovich, 2004). Studies that have applied memory load manipulations to a perspective-taking task have shown impaired ability to adopt a different perspective (Davis et al., 1996; Lin et al., 2010; Roxβnagel, C., 2000). Roxβnagel (2000) found that speakers who were put under cognitive load were less able to take the listener's need for information into account. Additionally, in language comprehension, an initially egocentric perspective was found with a delayed effect of correction (Keysar et al., 2000; Keysar & Barr, 2002). When people comprehend, they interpret what the speaker says from their own perspective with little consideration for the perspective of the speaker (Keysar et al., 2000). These findings suggest that overcoming egocentric bias requires effortful attention and requires working memory capacity.

In vehicle-pedestrian interaction, the driver can take a perspective when emitting a cue that is informative from the pedestrian's point of view, or the pedestrian can take perspective when inferring the intentions of the driver. The ability to take another perspective is an important component in vehicle-pedestrian interaction; it enables the pedestrian or driver to acquire information about the intention of the other and to predict their behavior, which results in safer and more efficient interaction (Palmeiro et al., 2018; Turnwald, Althoff, Wollherr & Buss 2016). Normally, when pedestrians make a crossing decision, they make use of vehicle dynamics such as vehicle speed and gap distance (Beggiato, Witzlack, & Springer, 2018; Dey & Terken, 2016; Zimmermann & Wettach, 2017) and implicit (eye contact, yielding) and explicit cues (hand gestures, honking, flashing lights) of the driver (Färber, 2016). For example, following the eye gaze of a driver may allow the pedestrian to understand the perception of the driver (e.g., knowing whether the driver has seen the pedestrian) and anticipate the future behavior of the driver.

1.2. External Human-Machine Interfaces (eHMIs)

To compensate for the lack of interpersonal communication between AVs and pedestrians, external human-machine interfaces (eHMIs) have been introduced (Lagstrom & Lundgren (2015). EHMIs appear in several forms, including text and symbolic messages, lights, and projections, and are intended to replace the cues that the driver communicates to pedestrians who intend to cross. Since these are novel concepts in vehicle-pedestrian interaction, they bring various issues and design challenges with them, which have yet to overcome. Besides, there must be no ambiguity regarding the message perspective of the AV. For example, a green brake light in front of the car could mean several things; the participant might think that the brake light refers to themselves (egocentric perspective) or the vehicle (allocentric perspective). If the pedestrian interprets the message as egocentric and therefore assumes it is safe to cross while the message was, in fact, designed to communicate that the vehicle was not yielding, hazardous situations may occur.

In autonomous vehicle-pedestrian (AV-P) interaction, the perspective taken by the pedestrian can be influenced by the design of the eHMI. From a pedestrian's perspective, the information can either be presented egocentrically (an advisory message) or allocentrically (an informatory message) (Dietrich, Willrodt, Wagner, & Bengler, 2018). 'Allocentric messages' refer to the intention of the vehicle itself; for instance, showing vehicle speed on an LED display to give the pedestrian information about the state (and possibly the intention) of the vehicle (Clamann, Aubert, & Cummings, 2017). 'Egocentric messages' refer to the pedestrian, presenting the text 'WALK' or 'DON'T WALK' on an LED display to inform the pedestrian that it is safe to cross (Dietrich et al., 2018).

It is presently unknown whether an eHMI should use an egocentric message perspective (e.g., Textual: 'Go ahead' (Ackermann, Beggiato, Schubert, & Krems, 2019), 'walk'/'don't walk' (Bazilinskyy, Dodou, & De Winter, 2019; De Clercq, Dietrich, Núñez Velasco, De Winter, & Happee, 2019; Kooijman et al., 2019; Fridman et al., 2017; Qin, 2019), 'CROSS NOW' (Matthews & Chowdhary, 2017), Symbolic: walking pedestrian silhouette (Deb, Strawderman, & Carruth, 2018; Fridman et al., 2017; Hudson, Deb, Carruth, McGinley, & Frey, 2018) , stop sign (Deb et al., 2018), actuated hand (Fridman et al., 2017) and Light strips: knightrider (De Clercq et al., 2019, Othersen et al., 2018) or an allocentric perspective (Textual: 'after you' (Nissan, 2015), 'braking' (Deb et al., 2018), stopping (Nissan, 2015), Symbolic: eyes on the car (Chang, 2017), Icon (indicating vehicle stops) (Weber et al., 2019) and Light strips: LED columns (Bockle, 2017); LED light strips (Habibovic et al., 2018; Petzoldt, Schleinitz, & Banse, 2018; Zhang, Vinkhuyzen, & Cefkin, 2018). Thus far, different conclusions regarding the use of message perspective have been reached (Reiff, De Winter, & Eisma, 2020). On the one hand, egocentric messages have been found to be more explicit and less ambiguous and result in higher perceived safety compared to allocentric messages as shown by computer simulation surveys (Ackermann et al., 2019; De Clercq et al., 2019; Fridman et al., 2017; Qin, 2019). They argue that explicit textual information such as ‘walk’ and ‘don't walk’ (see Table 2, De Clercq et al., 2019; Fridman et al., 2017) and familiar icons such as a walking pedestrian silhouette and upraised hand (Fridman et al., 2017) perform better than LED strips or allocentric textual and symbolic messages. Furthermore, explicit and egocentric gestures (from the pedestrian's point of view) are already used in traffic situations to resolve ambiguity (e.g., using a hand gesture to give right of way, traffic signs using a green walking pedestrian); therefore pedestrians are already familiar with these types of messages (McDougall, Curry, & De Bruijn, 1999). However, on the other hand, numerous studies indicate that the eHMI should give information about the state of the vehicle and that instructing pedestrians should be avoided (Bockle, 2017; Dietrich et al., 2018; Habibovic et al., 2018; Petzoldt et al., 2018; Zhang et al., 2018). Zhang et al. (2018) investigated the perception of pedestrians towards an AV with an LED light strip in a video-based experiment and found that participants interpreted the LED light as an allocentric message. In addition, instructing pedestrians might be dangerous in real traffic when multiple road users are present, the situation can become ambiguous, e.g., not knowing to whom the message is addressed (Dietrich et al., 2018).

Summarizing, perspective-taking allows humans to understand the perception of others and anticipate their future behavior. Whereas understanding what another person can see happens automatically, understanding how another person sees a particular stimulus is cognitively effortful. Furthermore, research in perspective-taking shows that humans rely on their own perspective and that switching to another perspective costs cognitive processing time. When designing an eHMI, the perspective taken by the pedestrian can be influenced by using an egocentric or allocentric perspective. However, there is no consensus on which perspective should be used in eHMI design.

1.3. Aim of the study

When designing an eHMI, it is essential to take into consideration that perspective may be affected by the characteristics of the eHMI as well as the pedestrians’ goals. Since little is known about the effect of message perspective on pedestrians crossing intentions, this study aims to investigate the effect of eHMI message perspective and cognitive load on participants’ perspective-taking, as inferred from their crossing intentions.

We designed a dual-task within-subjects experiment examining six different eHMI concepts. We used only textual messages since text appears to be the least ambiguous option among textual and symbolic displays and LED lights (Ackermann et al., 2019; De Clercq et al., 2019). To investigate the effect of message perspective, we selected the least ambiguous egocentric and allocentric messages from the literature. That is 'WALK' and 'DON’T WALK' (Bazilinskyy et al., 2019; De Clercq et al., 2019; Fridman et al., 2017; Hudson et al., 2018) for egocentric messages and 'BRAKING' (Deb et al., 2018) and 'DRIVING' for allocentric messages. To investigate the perception of pedestrians in AV-P interaction, we selected ‘GO’ and ‘STOP’, which was considered as ambiguous, and therefore could be interpreted either egocentric or allocentric by the pedestrian (Fridman et al., 2017). When a pedestrian needs to make a crossing decision at a crossroad, they need to integrate information from multiple sources (e.g., vehicles, cyclists, intersection, traffic signs) to form a mental representation of the situation. When the situation is more complex, more information needs to be considered, which results in a higher visual load. Furthermore, a pedestrian might be distracted or in a hurry, which can also result in increased cognitive load (Strayer & Fisher, 2016). To make the experiment more realistic from a cognitive perspective, we added a memory task to increase the cognitive load during the experiment (Lin et al., 2010).

The experimental design was implemented in a photo-based experiment using an eye tracker. The type of simulation resembles a combination of previous photo/video-based experiments (Ackermann et al., 2019; De Clercq et al., 2019; Fridman et al., 2017) and a visual perspective-taking (VPT) task the study of Todd et al. (2017). Participants were asked to press shift (L-shift – yes, R-shift is no) whether they could cross or not. Consequently, we examined the effect of message perspective and cognitive load, measuring response time and pupil size. Measurements of eye movement and reaction times have often been used to examine perspective-taking. Elevated gaze durations and increased reaction times when adjusting from a self to a different perspective have been used to infer egocentric bias and an elevated cognitive effort (Ferguson, Apperly, & Cane, 2017; Qureshi, Apperly, & Samson, 2010; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Furthermore, we measured crossing intentions and perceived clarity to investigate the clarity of the designs and which perspective pedestrians took when interpreting the eHMI messages.

Summarizing, we aimed to investigate the effect of eHMI message perspective and cognitive load on participants' perspective-taking, as inferred from their crossing intentions. The following research question is proposed:

What is the effect of message perspective, presented by external Human-Machine interfaces, on pedestrians' intentions, perspective taken and cognitive processes when making a crossing decision, as measured with various levels of cognitive load?

The effect of message perspective

H1. Egocentric messages will be regarded as clearer and less ambiguous compared to allocentric messages. This will be objectively reflected by more uniform crossing intentions and subjectively by a higher clarity rating.

H2. Crossing intentions will be cognitively less effortful when interacting with an egocentric perspective compared to an allocentric perspective. This will be objectively reflected by faster response times and smaller pupil diameter.

H3. When participants are put under cognitive load, they will have more difficulty interpreting the meaning of the message when making a crossing decision (Qureshi et al., 2010). This will be objectively reflected by decreased clarity score, longer response times and increased pupil diameter, but with a higher effect size for allocentric messages.

Perception of pedestrians towards ambiguous eHMIs

H4. When the message perspective is ambiguous, the crossing intentions of pedestrians will be less uniform compared to an explicit message perspective. Since both perspectives could be taken, it is also expected that the interaction with an ambiguous message is more demanding, which will be objectively reflected by longer response times and increased pupil diameter.

H5. As people tend to be egocentrically biased when comprehending, participants will take an egocentric perspective when interpreting ambiguous messages

H6. For the ambiguous messages, the egocentric bias will increase. That is even more participants will interpret the message egocentrically.

2. METHODS

2.1. Participants

Hundred and sixty-five MSc students from the Delft University of Technology from the course Human-Machine systems participated in this study. Prior data screening removed any response times before the onset of the image (too early) and response times higher than 5000ms (too late). Participants who made five or more mistakes (responding too early or too late) during the experiment were excluded (N = 62). Therefore, our final sample consists of 103 participants (68 males and 35 females), aged between 21 and 29 years (M = 23.3, SD = 2.0). Informed consent was obtained from all participants, and the experiment was approved by the TU Delft Human Research Ethics Committee. All participants were tested individually and followed the same procedure.

2.2. Materials and equipment

Eye movements were recorded binocularly at a sampling rate of 2000 HZ using an SR-Research Eyelink 1000 Plus eye tracker. Participants were asked to place their head in the head support during the entire experiment. The stimuli were shown on a 24.5-inch BENQ monitor with a resolution of 1920 x 1080 pixels (531 x 298 mm). The distance between the monitor and the table edge was 94 cm.

Figure 1. The experimental setup

2.3. Independent variables

Six eHMI concepts (see Figure. 3) were presented to the participants. The eHMIs were generated with the online tool LCD Display Generator (Avtanski, 2020). The eHMIs were all the same size and had white letters. We opted for white letters instead of color to prevent possible associations with traffic. It has been argued that people associate the color green with a moving vehicle and the color red with a stopping vehicle (Bazilinskyy et al., 2019; Zhang et al., 2018). Furthermore, although cyan is recommended to use for AVs because of its good visibility and because it is not yet used in traffic, this color might be interpreted as green (Bazilinskyy et al., 2019). We used a photo that was made during an earlier study for an experiment of Rodriguez Palmeiro et al. (2018). The concepts were placed on the bumper of a vehicle with a driver (Figure 2). We included an attentive driver instead of a distracted driver or no driver behind the wheel to avoid confusion on behalf of the participant. Pedestrians tend to be unwilling to cross when encountering an AV with an inattentive 'driver' or no driver behind the wheel at all (Malmsten Lundgren et al., 2017). Besides, only AVs at the most advanced level of automation (SAE level 5) will drive without human intervention or oversight, while for highly automated vehicles (SAE levels 3 and 4), human override is still required (SAE International, 2014).

Figure 2. One of the six visual stimuli used in the experiment: eHMI concept 'WALK'. The person in the driver seat provided written consent for the publication of this photo.

During the experiment, three independent variables were used. The first independent variable was the message perspective of the eHMI: (1) egocentric, (2) allocentric, or (3) ambiguous. Egocentric eHMIs indicate that the message is addressed to the pedestrian (i.e., giving advice). Allocentric eHMIs indicated that the message provides information about the intention of the vehicle. For ambiguous eHMIs, the message perspective is unclear in the sense that the message can be interpreted in two ways: egocentric or allocentric.

For the egocentric and allocentric eHMIs, the second independent variable is yielding behavior, namely whether the vehicle is yielding or non-yielding. The ambiguous messages are designed in such a way that it is open for interpretation; the eHMI could indicate both yielding or non-yielding behavior. Therefore, for the ambiguous eHMIs, yielding could not be defined yet (yielding behavior depends on the interpretation of the participants).

The third variable is the memory task: (1) baseline (2) low memory load and (3) high memory load. For the baseline condition, no memory task was added; in the low load memory task, the participant had to remember two digits, and in the high load memory task, the participant had to remember five digits.

Figure 3. The eHMI concepts used during the experiment. Left column: Egocentric messages, Middle column: Allocentric messages, Right column: Ambiguous messages

2.4. Dependent variables

2.4.1. Keypress task

The experiment consisted of a keypress task and a memory task. The keypress task was based on the procedure of Ferguson et al. (2017, exp 1). The primary stimuli consisted of a vehicle with one of the six eHMI concepts on the bumper (Figure 2). Before each primary stimulus, participants were shown the following statement: “I can cross”, prompting the participant to indicate whether they could cross or not. When the primary stimulus was shown, the participant had to respond using keys R-Shift for ‘yes’ and L-Shift for ‘no’. The participants were asked to respond as quickly as possible and had a maximum of 5 seconds to respond.

2.4.2. Memory task

For the memory task, we used a forward digit span task. The load of the memory task varied between 0 digits (baseline), 2 digits (low load), and 5 digits (high load). We chose for a maximum of 5 digits, based on Miller’s law, which argues that the number of objects humans can hold in short-term memory is 7 ± 2 (Miller, 1956). The digits were presented one by one before the onset of the statement ‘I can cross’. After responding to the primary stimulus, participants were asked to type in the digits they had remembered (Qureshi et al., 2010). During the baseline condition, participants had to type in 0 after they responded to the primary stimulus.

2.5. Procedure

The experiment consisted of 18 trials: six trials without memory task, six trials with a memory load of 2 digits, and six trials with a memory load of 5 digits. Each participant encountered the same eHMI with the same digits (See Appendix 4 for an overview of all stimuli, including their sequence of digits). The order in which the eHMIs were presented was random for each participant. After finishing all 18 trials, the participant was asked to rate the clarity of the six eHMI concepts on a scale of 1 to 10 (Appendix 5).

The trial began with a fixation cross in the center of the screen for 5750 ms when no memory task was included, followed by a blank screen for 1000 ms. When a memory task was included, the fixation cross was shown for 3750 ms for the low load memory task or 750 ms for the high load memory task. The fixation cross was replaced by a blank screen, which was shown for 250 ms, followed by a digit for 750 ms. This was repeated twice for the low load memory task and five times for the high load memory task. After the last digit was shown, a blank screen appeared for 1000 ms. After the blank screen, the statement ‘I can cross’ was shown for 2000 ms. Again, a blank screen was shown for 250 ms, followed by a fixation cross for 750 ms. Another blank screen was shown for 250 ms, which was replaced by the image showing the eHMI until the spacebar is pressed with a maximum duration of 5000 ms. Finally, on the last screen, the participant filled in the numbers s/he had to remember before. Figure 4 illustrates the sequence of stimuli of a trial.

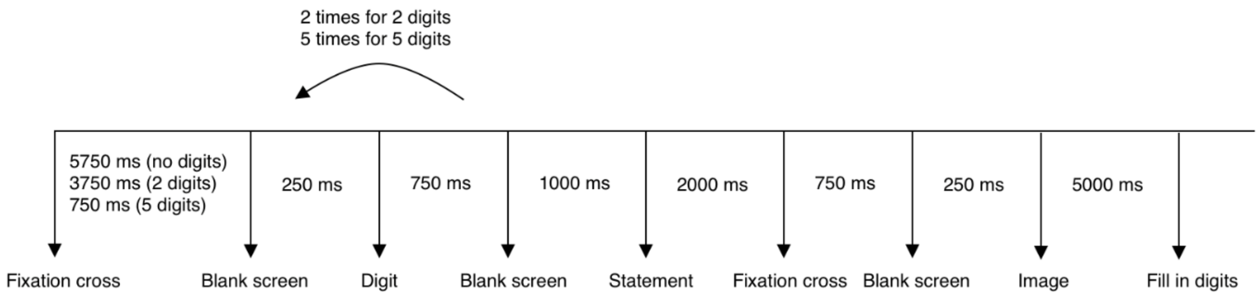


Figure 4. The sequence of a trial. Note that the time between the onset of the first fixation cross and the onset of the image was identical for each of the 18 trials.

2.6. Participants' task

Before the participant started the experiment, he/she read and signed the informed consent form (Appendix 1). Participants were presented with an introductory text informing them that they were about to view images of an automated vehicle with textual messages on the bumper and had to respond whether they could cross based on the displayed image. Furthermore, they were informed about the memory task and explained they needed to remember the digits until after they had responded to the image of the car. After the calibration was done, one practice trial was performed before starting the experiment (See Appendix 3 for steps and instructions). For the practice trial, we used a different message ('WILL STOP') to avoid familiarization. After the participant had completed all 18 trials, the six images were shown one by one, and the participant was asked to rate the clarity on a scale of 0 (completely agree) to 10 (completely disagree).

2.7. Dependent variables

A total of four subjective and subjective variables were analyzed. The subjective variable was self-reported clarity. The objective dependent variables were (1) response behavior (percentage pressing 'yes' and 'no') (2) response time and (3) pupil diameter.

2.7.1. Self-reported clarity

After the trials, the participants rated the clarity of the eHMI designs on a scale from 0 (completely disagree) to 10 (completely agree).

2.7.2. Response Behavior

To measure the crossing intentions of the participants, we analyzed the response behavior of the participants for each trial during the keypress task. For yielding vehicles ('WALK', 'BRAKING'), we calculated the percentage pressing 'yes' and for non-yielding vehicles ('DON'T WALK, 'BRAKING') we calculated the percentage pressing 'no' based on the answers of all participants that were included in the data processing. For the ambiguous messages ('GO', 'STOP') we analyzed the percentage pressing 'yes' or 'no', and based on the highest percentage pressing 'yes' or 'no' for each eHMI the message was placed under 'yielding' or 'non-yielding' behavior for further statistical analysis.

2.7.2.1. Clarity score

We used the percentage pressing 'yes' as an index of the clarity score. The clarity score was calculated as follows:

$$\text{Clarity score } (\%) = 2 * (\,|\text{ percentage pressing } \text{'yes'} - 50\%\,|\,)$$

A score of 100% resembles very clear, indicating that all participants interpreted the message the same way. A score of 0% resembles very ambiguous, indicating that 50% of the participants interpreted the message as they could cross the street and 50% as they could not cross the street.

2.7.3. Response time

Response time was measured since the moment the image of the eHMI was presented until the participant pressed the key 'L-shift' for no or 'R-shift' for yes. Response times were obtained for each trial and averaged for each memory task condition.

2.7.4. Pupil diameter

We extracted the participants' pupil diameter from the Eyelink eye-tracker data. Kahneman and Beatty (1966) demonstrated a relationship between pupil diameter and task complexity, and showed that increased task difficulty leads to dilation of the pupil. We used pupil diameter as an index of cognitive load.

2.8. Statistical analysis

First, a check of the responses of the 165 participants was performed. Apparently, when a participant pressed a key (any) before the onset of the stimulus, the participant already moved on to the next trial, meaning that participants were able to already give an answer before seeing the actual stimulus. Besides, if there was no response within 5 seconds, the participants automatically moved on to the next trial. Responses before the onset of the image (too early) and responses longer than 5000 ms after the onset of the image were considered as missing values. Furthermore, when participants did not use the appropriate keys (L-Shift or R-Shift), the response of the participant was not measured, and thus was saved as a missing value. Furthermore, participants had the option to not answer the questionnaire after the 18 trials, by pressing 'enter' without rating the eHMI design (see Appendix 5 for an overview of the questionnaire).

The self-reported clarity, response behavior (pressing 'yes' and 'no'), response time, and pupil diameter were analyzed by means of the mean and standard deviation. For pupil diameter, the mean was calculated by averaging the obtained data per 100 ms for each participant.

For yielding and non-yielding vehicles, a two-way full factorial repeated-measures analysis of variance was conducted to compare the main effects of the independent variables 'message perspective and memory task and the interaction between message perspective and memory task on response behavior, response time and pupil diameter. The results of the response behavior were binair values, encoded as 1 for pressing 'yes' and 0 for pressing 'no'. When using categorical variables as an indicator matrix, ANOVA is simply a special case of regression analysis. Furthermore, Hellevick (2009) has shown that the use of a linear regression and logistic have nearly identical outcomes when the variable is diochotomous. Since a linear regression can be safely used instead of a logistic regression, whereas ANOVA and linear regression are equivalent, ANOVA can be safely to analyze the repsonse behavior. Message perspective consisted of three levels (egocentric, allocentric, and ambiguous), and memory task consisted of three levels (no memory task (baseline), low load (2 digits), and high load (5 digits)). Furthermore, we performed a one-way repeated measures analysis of variance to measure the effect of memory task on the pupil diameter for the whole trial.

For the self-assessed clarity rating, a one-way full factorial repeated-measures analysis of variance was conducted to analyze the effect of each eHMI concept.

Significant differences between the conditions were assessed with MATLAB's *multcompare* function, using the Bonferroni critical value.

3. RESULTS

3.1. Data quality assessment

Prior data screening removed any response times before the onset of the image (too early) and response times of > 5000 ms (too late). Furthermore, per trial, we examined whether eye-tracking data was missing. Participants who had five or more invalid trials (responded too early, too late, did not respond or had missing eye-tracking data) were excluded (N = 62).

3.2. Self-assessed clarity

After conducting the experiment, the participants were asked to indicate the clarity of the eHMI on a scale from 1-10 while showing the photo of the six eHMI designs once again. The results show that the egocentric messages were regarded as most clear. The message 'DON'T WALK received the highest clarity rating, followed by 'WALK', 'DRIVING, 'BRAKING', 'GO', and 'STOP' (Table 1). One participant did not rate any designs. Furthermore, for the message 'GO', 5 participants did not provide an answer, and for the message 'STOP', 12 participants did not rate the message.

A one-way full-factorial repeated-measures ANOVA was conducted to compare the effect of message perspective on perceived clarity. There was a significant effect in clarity rating for the different eHMI concepts, $F(5,440) = 19.535$ $p < 0.001$, $\eta_p^2 = 0.182$. Post hoc comparisons using the Bonferroni correction showed that the egocentric messages 'WALK' and 'DON'T WALK were rated significantly higher compared to allocentric and ambiguous messages (Table 1). Furthermore, the message 'STOP' received a significantly lower clarity rating compared to the clarity relative to all messages, revealing that the message 'STOP' was regarded as most ambiguous. No differences in clarity ratings were found between 'BRAKING', 'DRIVING', and 'GO'.

Table 1. Overview of the post hoc comparisons using the Bonferroni correction for clarity rating. The critical p-value was 0.05. Bold values indicate significant p-values.

(I) eHMI	*(J) eHMI*	*Mean (SD)*	*Mean (SD)*	*p-value*
WALK	DON'T WALK	8.82 (1.88)	9.13 (1.79)	0.667
WALK	BRAKING	8.82 (1.88)	7.26 (2.37)	**< 0.001**
WALK	DRIVING	8.82 (1.88)	7.68 (2.32)	**0.010**
WALK	GO	8.82 (1.88)	7.07 (2.68)	**< 0.001**
WALK	STOP	8.82 (1.88)	6.58 (2.96)	**< 0.001**
DON'T WALK	BRAKING	9.13 (1.79)	7.26 (2.37)	**< 0.001**
DON'T WALK	DRIVING	9.13 (1.79)	7.68 (2.32)	**< 0.001**
DON'T WALK	GO	9.13 (1.79)	7.07 (2.68)	**< 0.001**
DON'T WALK	STOP	9.13 (1.79)	6.58 (2.96)	**< 0.001**
BRAKING	DRIVING	7.26 (2.37)	7.68 (2.32)	0.401
BRAKING	GO	7.26 (2.37)	7.07 (2.68)	1.000
BRAKING	STOP	7.26 (2.37)	6.58 (2.96)	1.000
DRIVING	GO	7.68 (2.32)	7.07 (2.68)	1.000
DRIVING	STOP	7.68 (2.32)	6.58 (2.96)	**0.035**
GO	STOP	7.07 (2.68)	6.58 (2.96)	**< 0.001**

3.3. Response behavior

Figure 5 shows the distribution of responses for each stimulus. The majority of the participants interpreted the message ‘WALK’ and ‘BRAKING’ as safe to cross and the message ‘DON’T WALK’ and ‘DRIVING’ as not safe to cross, which matches the intention of the designs. Regarding the perspective taken for the ambiguous messages, the figure shows that most respondents interpreted the messages as egocentric. For the message ‘GO’, the majority pressed ‘yes’, indicating that the respondents associated the message ‘GO’ as permission to cross. For the message ‘STOP’ the majority pressed ‘no’, indicating that respondents associated the message ‘STOP’ as an indicator not to cross. Therefore, the message ‘GO’ will be regarded as a yielding vehicle and the message ‘STOP’ as a non-yielding vehicle for statistical analysis. Contrary to our expecations, memory task scarcely affected the crossing intentions of the participants, as can be clearly seen in Figure 5 and Table 3.

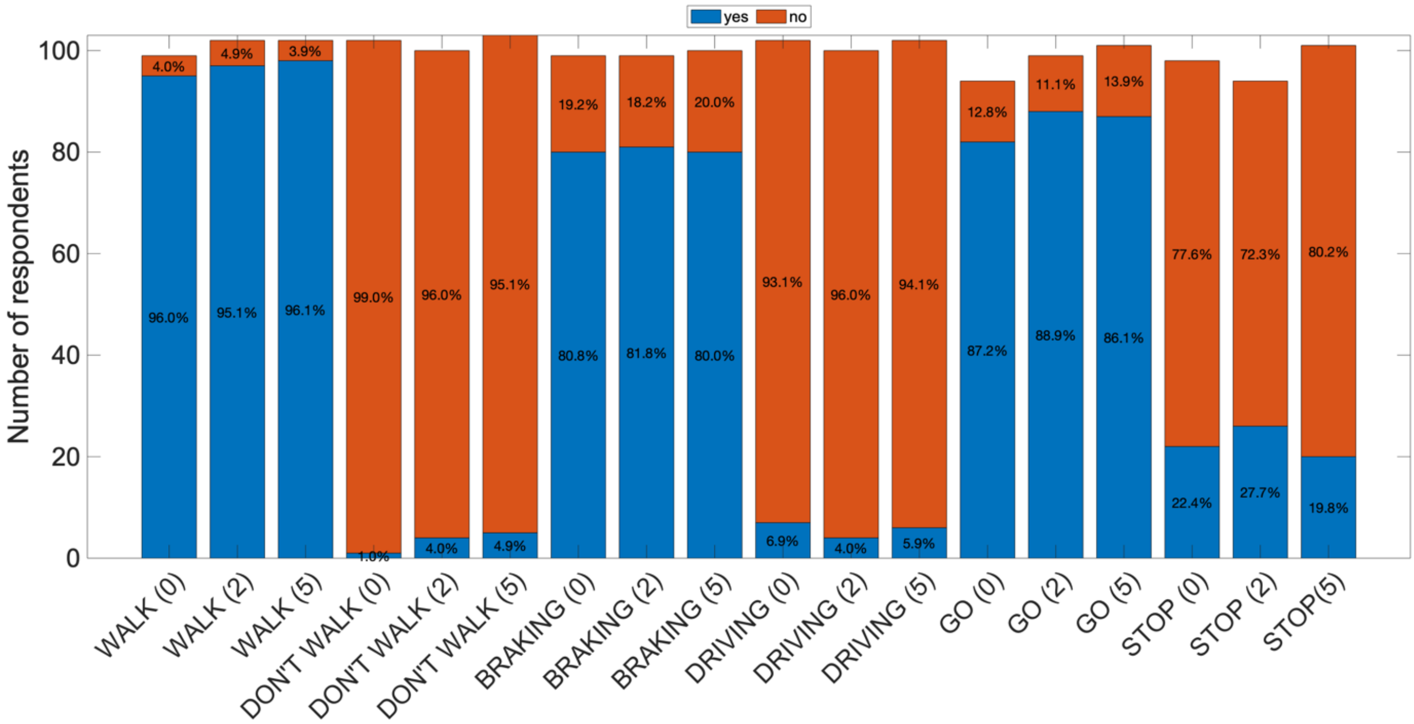


Figure 5. Distribution of answers to the statement 'I can cross' for each stimulus. The number between parentheses indicates the number of digits during the memory task. 0 stands for no memory task, 2 stands for the low load memory task and 5 stands for the high load memory task. The percentage is calculated based on the number of respondents for each stimulus.

Based on the responses from the keypress task, the clarity score was calculated for each stimulus (Table 2). For yielding vehicles, the effect of message perspective was in the expected direction; the egocentric ‘WALK’ resulted in a higher clarity score than the allocentric ‘BRAKING’ and ambiguous ‘GO’ (Table 2). On the contrary, for non-yielding vehicles, the effect of message perspective was less evident; no difference in clarity score was measured between the egocentric ‘DON’T WALK’ and the allocentric ‘DRIVING’, whereas the ambiguous ‘STOP’ resulted in a much lower clarity score (Table 2). These results indicate that the crossing intentions were most uniform for the egocentric perspective, and thus resulted in a higher clarity compared to an allocentric or ambiguous perspective. The clarity score as a function of the trials is shown in Appendix 7. The figure provides illustrative learning curves, which were fit using the following function: y = a*exp(b*x)+c. No learning effect was found for the message ‘DON’T WALK, whereas a diminishing learning effect can be distinguished for the egocentric ‘WALK’, the allocentric ‘BRAKING’ and ‘DRIVING’ and the ambiguous ‘GO’, arguing that the maximum clarity score was achieved during the experiment. For the message ‘STOP’ it was found that the curve is still increasing at the end of the experiment, indicating that participans were still updating their responses. Table 2 shows the clarity score averaged over the last six trials (e.g., maximum clarity score) for each message, showing the greatest increase in clarity score for the ambiguous message in comparison to the clarity score averaged over the whole experiment. As for the other messages, the clarity score remained roughly consistent.

According to a two-way full factorial repeated measures ANOVA, message perspective was significant, whereas memory task and the interaction between message perspective and memory task were not significant for yielding vehicles (Table 3). Post hoc comparisons using the Bonferroni correction showed that the crossing intentions (e.g., percentage pressing 'yes') were significantly more uniform for the egocentric 'WALK' compared to the allocentric 'BRAKING' and ambiguous 'GO'. (Table 4). No significant difference was found between 'BRAKING' and 'GO'. Interestingly, even though 'BRAKING' was designed without ambiguity and 'GO' was designed to be ambiguous, the ambiguous 'GO' resulted in more uniform crossing intentions, as evidenced by a higher percentage pressing 'yes'. For non-yielding vehicles, message perspective, as well as the interaction between message perspective and memory task, were significant, whereas memory task was not significant (Table 3). Post hoc comparisons showed that during the baseline condition (e.g., memory task), the egocentric 'DON'T WALK' resulted in a higher percentage intending not to cross compared to the allocentric 'DRIVING' and ambiguous 'STOP' (Table 4). Altough we expected that the the crossing intentions would be less consistent when performing a concurrent memory task, especially for allocentric messages, an even higher percentage pressed 'no' for the allocentric 'DRIVING' compared to the baseline condition. For both memory load conditions (e.g., low and high), no difference was found for 'DON'T WALK' and 'DRIVING', whereas 'STOP' resulted in a significantly lower percentage pressing 'no' (except to 'DRIVING' during the high load memory task) (Table 4). Between yielding and non-yielding vehicles, the effect size of message perspective on pedestrians crossing intentions is about twice as strong for non-yielding vehicles (η_p^2=0.148) as the effect size for yielding vehicles (η_p^2=0.078)

Table 2. Overview of the clarity scores. Based on the responses from the keypress task, the clarity score is calculated with the formula from Section 2.7.4. The overall mean present mean of each message averaged over the three memory load conditions. The mean clarity score for learning behavior presents the averaged clarity score over the last six trials of the experiment.

EHMI	*Overall mean*	*Baseline*	*Low load*	*High load*	*Mean (Learning behavior)*
WALK	91.4%	91.9%	90.2%	92.2%	93.8%
DON'T WALK	93.4%	98.0%	92.0%	90.3%	93.3%
BRAKING	61.8%	61.6%	63.6%	60.0%	61.8%
DRIVING	88.8%	86.3%	92.0%	88.2%	94.7%
GO	74.8%	74.5%	77.8%	72.3%	85.0%
STOP	53.4%	55.1%	44.7%	60.4%	74.3%

Table 3. Overview of repeated ANOVA measures of percentage pressing 'yes' for yielding vehicles and 'no' for non-yielding vehicles for message perspective, memory load and interaction between message perspective and memory load. The critical p-value was 0.05. Bold values indicate significant p-values.

Variable	*DF1*	*DF2*	*F*	*p-value*	η_p^2
Yielding					
Message perspective	2	156	6.599	**< 0.001**	0.078
Memory task	2	156	0.430	0.362	0.013
Perspective * task	4	312	1.035	0.389	0.013
Non-yielding					
Message perspective	2	156	13.573	**< 0.001**	0.148
Memory task	2	156	0.430	0.651	0.006
Perspective * task	4	312	3.054	**0.017**	0.038

Abbreviations: DF = Degrees of Freedom, Perspective * task = message perspective * memory task

Table 4. Overview of the post hoc comparisons using the Bonferroni correction with p-value and the mean (SD) percentage pressing 'yes' for yielding vehicles and pressing 'no' for non-yielding vehicles of each group within the tested variable. The critical p-value was 0.05. Bold values indicate significant p-values.

(I) Perspective	*(J) Perspective*	*Mean (SD)*	*Mean (SD)*	*p-value*
Yielding		Percentage pressing 'yes'		
Egocentric (WALK)	Allocentric (BRAKING)	95.7 (20.3)	80.9 (39.4)	**0.003**
Egocentric (WALK)	Ambiguous (GO)	95.7 (20.3)	87.3 (33.2)	**0.032**
Allocentric (BRAKING)	Ambiguous (GO)	80.9 (39.4)	87.3 (33.2)	0.421
Non-yielding		Percentage pressing 'no'		
Egocentric (DON'T WALK)	Allocentric (DRIVING)	96.7 (17.8)	94.4 (23.0)	0.188
Egocentric (DON'T WALK)	Ambiguous (STOP)	96.7 (17.8)	76.8 (42.3)	**< 0.001**
Allocentric (DRIVING)	Ambiguous (STOP)	94.4 (23.0)	76.8 (42.3)	**0.009**
(I) Memory task	*(J) Memory task*	*Mean (SD)*	*Mean (SD)*	*p-value*
Yielding		Percentage pressing 'yes'		
Baseline	Low load	88.0 (32.5)	88.7 (31.8)	1.000
Baseline	High load	88.0 (32.5)	87.5 (33.2)	0.717
Low load	High load	88.7 (31.8)	87.5 (33.2)	0.669
Non-Yielding		Percentage pressing 'no'		
Baseline	Low load	90.1 (30.0)	88.4 (32.0)	1.000
Baseline	High load	90.1 (30.0)	89.9 (30.2)	1.000
Low load	High load	88.4 (32.0)	89.9 (30.2)	1.000

Table 5. Overview of the post hoc comparisons using the Bonferroni correction among the interaction between message perspective and memory task for percentage pressing 'no' for non-yielding vehicles. The critical p-value was 0.05. Bold values indicate significant p-values.

Memory task	*(I) Perspective*	*(J) Perspective*	*p-value*
Baseline	Egocentric (DON'T WALK)	Allocentric (DRIVING)	**0.040**
	Egocentric (DON'T WALK)	Ambiguous (STOP)	**< 0.001**
	Allocentric (DRIVING)	Ambiguous (STOP)	0.177
Low load	Egocentric (DON'T WALK)	Allocentric (DRIVING)	0.961
	Egocentric (DON'T WALK)	Ambiguous (STOP)	**< 0.001**
	Allocentric (DRIVING)	Ambiguous (STOP)	**< 0.001**
High load	Egocentric (DON'T WALK)	Allocentric (DRIVING)	0.544
	Egocentric (DON'T WALK)	Ambiguous (STOP)	**0.007**
	Allocentric (DRIVING)	Ambiguous (STOP)	0.114

3.4. Response times

Figure 6 shows the mean response times for each stimulus. The fastest response times averaged over memory task were found for the message 'GO', followed by 'DRIVING', 'DON'T WALK', 'WALK', 'BRAKING' and the ambiguous 'STOP' (Table 8). As for the baseline condition, fastest response times were found for egocentric messages compared to allocentric and ambiguous messages. The intermediate response times were found for the allocentric 'DRIVING' and ambiguous 'GO', and the longest response times were found for the allocentric 'BRAKING' and ambiguous 'STOP' (Table 6). Furthermore, the figure shows that memory task had a substantial effect on response times. Contrary to our expectations, the presence of a memory task reduced the amount of time to make a crossing decision compared to no memory task (e.g., baseline) as evidenced by the faster response times. For the egocentric 'WALK' and 'DON'T WALK, the allocentric 'DRIVING' and the ambiguous 'GO', practically no difference was found between the different loads, while for the allocentric 'BRAKING' and the ambiguous 'STOP' only the high load memory task affected the response times (Figure 6, Table 6). The effect of memory task was highest for 'DRIVING' and 'GO'. The response time decreased respectively 16.5% and 16.2% (high load), against respectively 6.0% and 5.2% (high load) for 'WALK' and 'DON'T WALK' and 5.2% and 6.1% (high load) for 'BRAKING' and 'STOP with respect to the baseline condition (see Appendix 11 for an overview of all combinations of conditions). The data shows even faster response times for the message 'DRIVING' and 'GO' compared to the egocentric messages and the message 'BRAKING' and 'STOP' (Table 6). See appendix 11 for an overview of the mean response times for each stimulus for each trial.

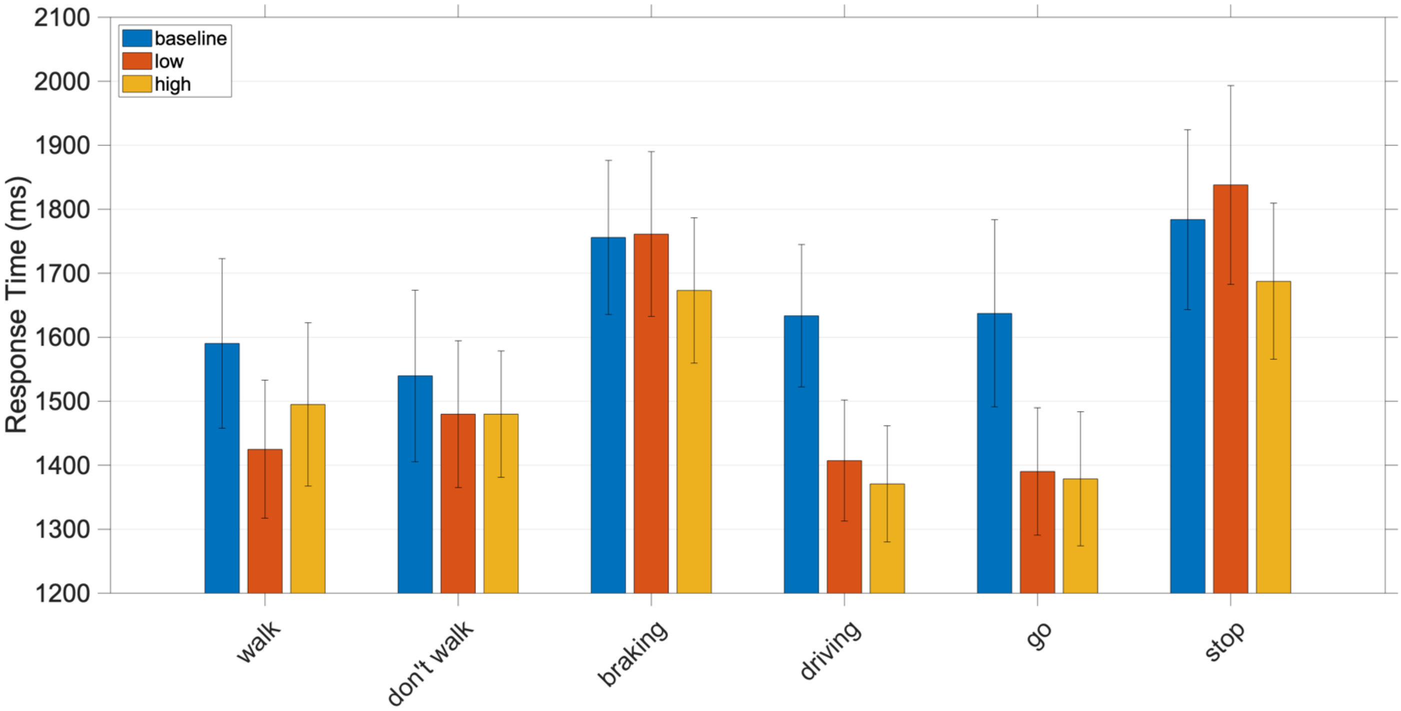


Figure 6. Response time (in milliseconds) was calculated for each trial by obtaining the mean response time for all participants who provided an answer. The response time was measured from the moment the image was onset until the spacebar was pressed. Error bars correspond to within-subjects 95% confidence intervals (Cousineau, 2005).

According to a two-way full-factorial repeated-measures ANOVA, the effect of message perspective, as well as the effect of memory task, was significant, whereas no significant interaction was found between message perspective and memory task for yielding vehicles (Table 7). For non-yielding vehicles, a significant difference was found for message perspective, whereas no significant difference was found for memory task as well as the interaction between message perspective and memory task (Table 7). The effect size of message perspective was stronger for non-yielding vehicles (η_p^2=0.217) than yielding vehicles (η_p^2=0.128). Post hoc comparisons showed that 'BRAKING' and 'STOP' resulted in significantly longer response times compared to the other messages, no significant difference was found between the other combinations (Table 8). In other words, respondents did not have more difficulty interpreting the allocentric 'DRIVING and the ambiguous 'GO' compared to egocentric messages in terms of cognitive processing time. A, a stronger effect size for message perspective was found for non-yielding vehicles higher effect size was found for non-yielding vehicles

As for the memory task, post hoc comparisons showed that both the low load and high load memory task resulted in faster response times compared to the baseline condition for yielding vehicles (Table 8). No differences were found between the low and high load memory task. For non-yielding vehicles, no significant differences were found for memory task. As can be seen in Figure 6, the effect of memory task was large for 'DRIVING', whereas the response times for 'DON'T WALK' and 'STOP' were less affected. Appendix 7 shows the effect of memory load on the learning behavior as a function of the trials. At the beginning of the experiment participants responded fastest for the high load condition, followed by the low load condition and longest response times were found during the baseline condition. Though, a ceiling effect can be distinguished around the 17th and 18th trial for response times, indicating that at the end of the experiment a saturation has been reached; the memory task had no longer an effect of the response times.

Table 6. Overview of the mean (SD) response times (ms) for yielding and non-yielding vehicles for each stimulus.

	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Yielding	*WALK*		*BRAKING*		*GO*	
Baseline	1590	808	1756	809	1637	940
Low	1425	679	1761	840	1390	632
High	1495	775	1673	754	1379	720
Non-yielding	*DON'T WALK*		*DRIVING*		*STOP*	
Baseline	1540	837	1634	790	1984	900
Low	1480	826	1407	565	1838	1031
High	1480	697	1371	525	1687	811

Table 7. Overview of repeated ANOVA measures of response times for message perspective, memory task, and message perspective and memory task for both yielding and non-yielding vehicles. The critical p-value was 0.05. Bold values indicate significant p-values.

Variable	*Df1*	*Df2*	*F*	*p-value*	η_p^2
Yielding					
Perspective	2	156	11.422	**< 0.001**	0.128
Memory task	2	156	4.517	**0.012**	0.055
Perspective * task	4	312	1.247	0.291	0.016
Non-yielding					
Perspective	2	156	21.614	**< 0.001**	0.217
Memory task	2	156	2.246	0.109	0.028
Perspective * task	4	312	1.223	0.301	0.015

Table 8. Overview of the post hoc comparisons using the Bonferroni correction with p-value and the mean (SD) response times (ms) of each group within the tested variable. For message perspective, the response times are averaged over all memory task conditions. For memory task, the response times for each memory load were averaged over the eHMIs. The critical p-value was 0.05. Bold values indicate significant p-values.

(I) Perspective	*(J) Perspective*	*Mean (SD)*	*Mean (SD)*	*p-value*
Yielding				
Egocentric (WALK)	Allocentric (BRAKING)	1503 (776)	1730 (800)	**0.015**
Egocentric (WALK)	Ambiguous (GO)	1503 (776)	1465 (778)	0.281
Allocentric (BRAKING)	Ambiguous (GO)	1730 (800)	1465 (778)	**< 0.001**
Non-yielding				
Egocentric (DON'T WALK)	Allocentric (DRIVING)	1500 (787)	1471 (646)	1.000
Egocentric (DON'T WALK)	Ambiguous (STOP)	1500 (787)	1768 (915)	**< 0.001**
Allocentric (DRIVING)	Ambiguous (STOP)	1471 (646)	1768 (915)	**< 0.001**
(I) Memory task	*(J) Memory task*	*Mean (SD)*	*Mean (SD)*	*p-value*
Yielding				
Baseline	Low load	1662 (853)	1524 (739)	**0.041**
Baseline	High load	1662 (853)	1515 (757)	**0.063**
Low load	High load	1524 (739)	1515 (757)	1.000
Non-Yielding				
Baseline	Low load	1651 (846)	1570 (843)	0.432
Baseline	High load	1651 (846)	1512 (698)	0.193
Low load	High load	1570 (843)	1512 (698)	1.000

Abbreviations: DF = Degrees of Freedom, Perspective * task = message perspective * memory task.

3.5. Correlation analysis

Figure 8 shows scatter plots of the self-reported clarity and the clarity score (left plot) and of the mean respone time and clarity score (right plot) for the baseline condition. It reveals a strong positive correlation (r = 0.91) between clarity rating and clarity score. An even stronger correlation was found between the clarity score and the response times (r = -0.97). In other words, it appears that clarity score and response times are both affected by the same mechanism, which we think the degree of explicitness of the message. The correlation between clarity score and response time decreased as the memory load increased, which is due to the fact that the crossing intentions remained consistent while the response times decreased (r = -0.903 for low load and r = -0.705 for high load).

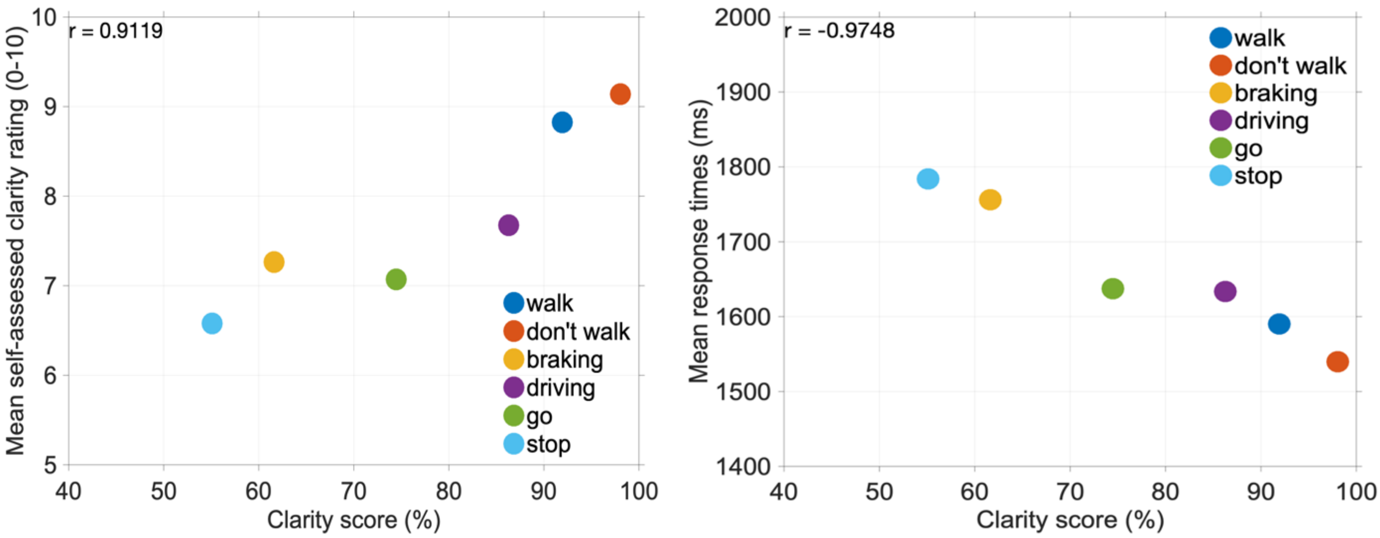


Figure 7. Left: Correlation between self-reported clarity and clarity score for the baseline condition. Right: Correlation between mean response time and clarity score for the baseline condition. *r* shows the Pearson correlation coefficient.

3.6. Pupil diameter

Figure 8 shows the pupil diameter for the three memory task conditions as a function of time. The pupil diameter for the high load memory task starts increasing around t = 1.75 s, which is right after the presentation of the first digit. As can be seen, the pupil diameter keeps increasing for each digit added to the memory task. For the low load memory task, the pupil diameter starts increasing around t = 4.5, which is when the first digit was presented. Furthermore, our analyses of pupil size over time shows an early peak at 7 seconds, which is the time the statement ‘I can cross’ was shown. According to a one-way full factorial ANOVA the effect of memory task was significant for most time bins (see Table 4). Post hoc comparisons showed that from T = 4, the pupil diameter started to become significantly bigger ($p < 0.001$) for the trials where participants performed the high load memory task compared to the baseline condition. For T_7 and T_8, the pupil diameter became significantly bigger ($p < 0.001$) when participants performed the low load memory task compared to the baseline condition (See Appendix 13 for all combinations).

When looking at the mean response time (vertical lines), no difference in pupil diameter was found for yielding vehicles whereas a small difference was found for non-yielding vehicles between egocentric and allocentric message (see Appendix 12). Furthermore, it can be seen that for both yielding and non-yielding vehicles the ambiguous messages resulted in the biggest pupil diameter. Interestingly, even though the fastest response times were found for the message ‘GO’, a bigger pupil diameter was found, indicating that ambiguity resulted in (slightly) higher cognitive load. Since we averaged the pupil diameters over 1000ms, we performed ANOVA for T_{10} (i.e., the onset of the image (t = 10 – 10.99)) and for T_{11} (i.e., 1 second after the onset of the image (t = 11 – 11.99)). According to a two-way full factorial ANOVA the effect of memory task was significant, whereas the effect of message perspective and the interaction between memory task and message perspective was not significant at T_{10} and T_{11} for yielding vehicles (see Table 6 & 7). For yielding vehicles, the effect of memory task was significant, whereas the interaction between message perspective and memory task was not significant for T_{10} and T_{11} (see Table 6 & 7). The effect of message perspective was not significant for T_{10} but was significant for T_{11}.

For both yielding and non-yielding vehicles, post hoc comparisons showed no difference in pupil diameter between the baseline condition and the low memory task, but an increase in pupil diameter was found when participants performed a high load memory task compared to a low load memory task ($p < 0.001$) and no memory task at all ($p < 0.001$) for T_{10} and T_{11}. However, as one can see, the pupil size was already bigger at the onset of the image for the high load memory task, whereas there was hardly a difference in pupil size at the averaged moment of response (Figure 8). Although, message perspective appeared to be significant for non-yielding vehicles according to ANOVA, post hoc comparisons did not show a significant difference. As can be seen in Table 7, the pupil diameter was smallest for the message 'DON'T WALK', followed by 'DRIVING' and largest diameter was found for the message 'STOP'; however, the differences were very small. In other words, message perspective did not have a substantial effect on the cognitive load (e.g., pupil diameter) of the respondents.

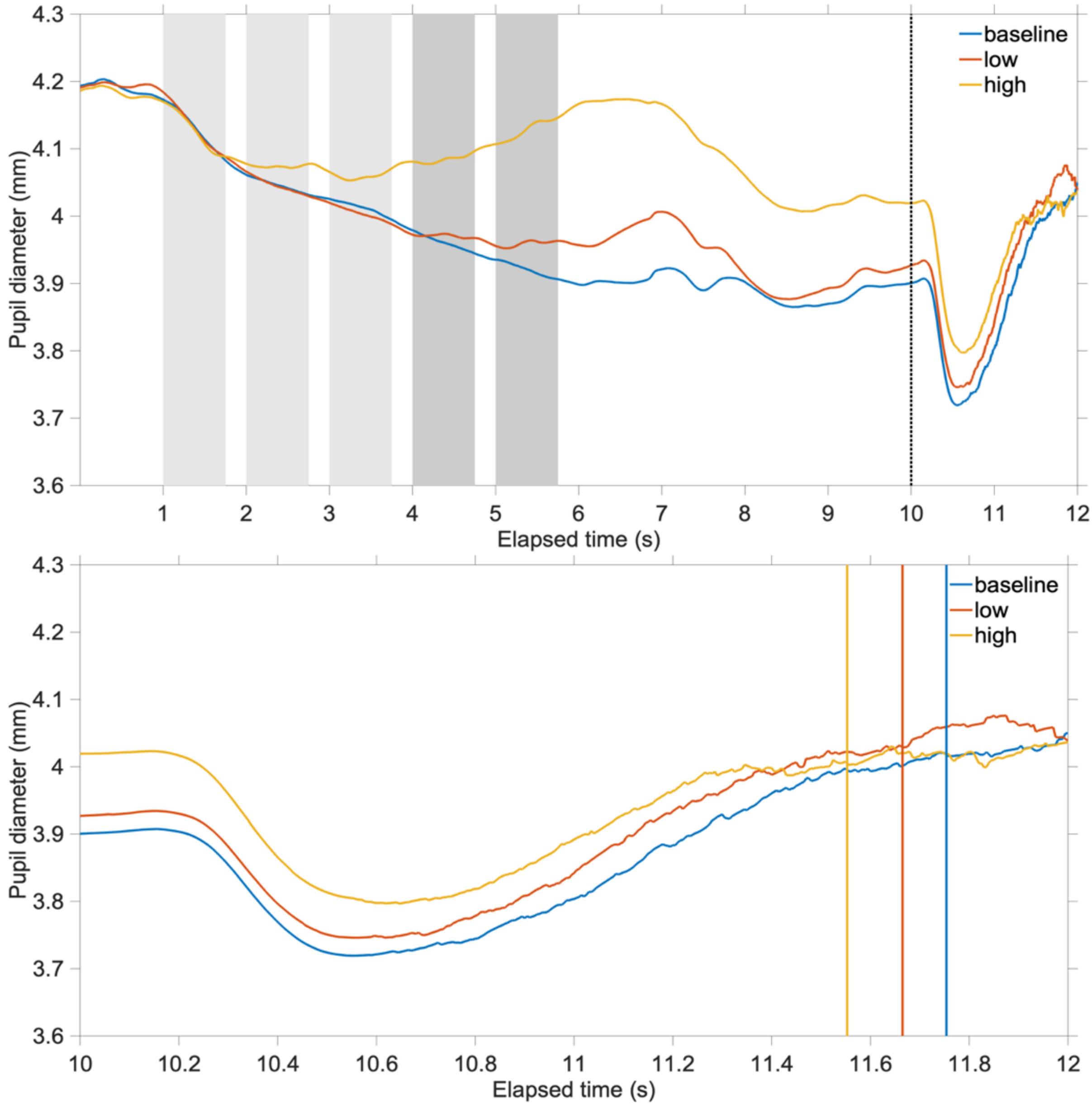


Figure 8. Mean pupil diameter for each memory task as a function of time. The grey vertical lines represent the onset of each digit and the black dotted vertical lines represent the onset of the stimulus. Pupil size was sampled every 2 ms. Pupillary responses are higher for increased cognitive load. Figure 7b. A close up of figure a starting at t=10, which is the onset of the figure. The colored vertical lines represent the average response times for each memory task. Baseline stands for the baseline condition, low stands for the low load memory task and high stands for the high load memory task.

Table 9. Overview of ANOVA measures for memory task effects. Mean pupil diameter per participant was averaged over every 1000ms. The critical p-value was 0.05. Bold values indicate significant p-values.

Time (sec)	*DF1*	*DF2*	*F*	*p-value*	η_P^2
T_1 (0 – 1)	2	134	0.213	0.788	0.002
T_2 (2 – 3)	2	134	0.129	0.991	0.001
T_3 (3 – 4)	2	134	7.01	**0.011**	0.069
T_4 (4– 5)	2	134	19.73	**<.001**	0.174
T_5 (5 – 6)	2	134	70.31	**<.001**	0.428
T_6 (6 – 7)	2	134	122.2	**<.001**	0.565
T_7 (7 – 8)	2	134	142.9	**<.001**	0.603
T_8 (8 – 9)	2	134	93.40	**<.001**	0.500
T_9 (9 – 10)	2	134	72.42	**<.001**	0.435
T_{10} (10 – 11)	2	134	70.80	**<.001**	0.430
T_{11} (11 – 12)	2	134	50.97	**<.001**	0.352
T_{12} (12 – 13)	2	22	6.189	0.126	0.198

Table 10. Overview of ANOVA measures for the first second of the onset of the image (T_{10} (t = 10 – 10.99)) and one second after the onset of the image (T_{11} t = 11-11.99). Mean pupil diameter per participant was averaged 1000ms for each trial. The critical p-value was 0.05. Bold values indicate significant p-values.

Variable	*Time (sec)*	*DF1*	*DF2*	*F*	*p-value*	η_P^2
Yielding						
Message perspective	T_{10} (10 – 11)	2	156	2.000	0.139	0.025
	T_{11} (11 – 12)	2	156	0.853	0.428	0.011
Memory task	T_{10} (10 – 11)	2	156	30.24	**<0.001**	0.279
	T_{11} (11 – 12)	2	156	25.16	**<0.001**	0.244
Perspective * task	T_{10} (10 – 11)	4	312	0.267	0.899	0.003
	T_{11} (11 – 12)	4	312	0.756	0.555	0.010
Non-yielding						
Message perspective	T_{10} (10 – 11)	2	156	2.202	0.114	0.028
	T_{11} (11 – 12)	2	156	3.125	**0.047**	0.039
Memory task	T_{10} (10 – 11)	2	156	37.06	**<0.001**	0.322
	T_{11} (11 – 12)	2	156	22.04	**<0.001**	0.220
Perspective * task	T_{10} (10 – 11)	4	312	0.609	0.657	0.008
	T_{11} (11 – 12)	4	312	0.756	0.555	0.010

Abbreviations: Perspective stands for message perspective, perspective * task stands for message perspective * memory task.

Table 11. Overview of the mean (SD) pupil diameter (mm) for yielding and non-yielding vehicles for T_{10} and T_{11}. For eHMI, the response times are averaged over all memory task conditions. For memory task, the response times are averaged over the eHMIs.

Variable		*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Yielding							
Message perspective		*WALK*		*BRAKING*		*GO*	
	T_{10}	3.931	0.472	3.945	0.472	3.926	0.469
	T_{11}	3.840	0.462	3.838	0.459	3.841	0.457
Memory task		*Baseline*		*Low*		*High*	
	T_{10}	3.886	0.447	3.904	0.478	4.010	0.477
	T_{11}	3.800	0.444	3.829	0.463	3.890	0.466
		WALK		*BRAKING*		*GO*	
Baseline	T_{10}	3.868	0.446	3.909	0.450	3.882	0.448
	T_{11}	3.787	0.441	3.811	0.444	3.799	0.435
Low	T_{10}	3.919	0.506	3.911	0.478	3.880	0.451
	T_{11}	3.840	0.481	3.825	0.461	3.820	0.450
High	T_{10}	4.005	0.455	4.014	0.483	4.011	0.498
	T_{11}	3.890	0.461	3.877	0.472	3.901	0.469
Non-yielding							
Message perspective		*DON'T WALK*		*DRIVING*		*STOP*	
	T_{10}	3.925	0.481	3.964	0.486	3.963	0.493
	T_{11}	3.827	0.470	3.852	0.465	3.880	0.385
Memory task		*Baseline*		*Low*		*High*	
	T_{10}	3.892	0.469	3.923	0.480	4.035	0.499
	T_{11}	3.808	0.472	3.839	0.477	3.910	0.466
		DON'T WALK		*DRIVING*		*STOP*	
Baseline	T_{10}	3.870	0.482	3.884	0.451	3.923	0.477
	T_{11}	3.798	0.497	3.788	0.435	3.849	0.485
Low	T_{10}	3.898	0.456	3.934	0.489	3.937	0.499
	T_{11}	3.825	0.438	3.834	0.486	3.880	0.509
High	T_{10}	4.007	0.497	4.074	0.500	4.025	0.503
	T_{11}	3.888	0.469	3.820	0.466	3.909	0.465

3.7. Memory task

The mean percentage correct responses for the memory task was 98.7% (SD = 11.1) for the baseline condition, 89.8% (SD = 30.3) for the low load memory task and 70.6% (SD = 45.6) for high load memory task. A two-way full -actorial repeated-measures ANOVA revealed significant differences in correctly performing the memory task between eHMIs, $F(2,204) = 71.54$, $p <.001$, $\eta_p^2 = 0.412$ and for the interaction between message perspective and memory task, $F(10,1020) = 2.220$, $p = 0.001$, $\eta_p^2 = 0.02$. When performing the high load memory task, more mistakes were made for the eHMI 'DRIVING' (M = 60%) than for the message 'DON'T WALK (M = 81%, p = 0.007) and 'BRAKING' (M = 80% p = 0.006). However, as mentioned in the method section, the digits for the memory task were fixed, indicating that each participant encountered the same sequence of digits for each stimulus. Therefore, we argue that the results of the memory task are not correlated with the stimulus, but rather have to do with the difficulty of the sequence of the digits. To summarize, the sequence of digits for the high load memory task was more difficult for the message 'DRIVING' compared to 'DON'T WALK' and 'BRAKING'.

3.8. Correlation analysis

Table 12 shows the Pearson correlation matrix for age, gender, and dependent measurements. The bold values in the table indicate significant values for $p < 0.05$ for testing the null hypothesis that the correlation is zero. The table shows that the pupil diameter was correlated with age, whereas a lower pupil diameter was found for younger participants. Furthermore, a correlation was found for gender and response time, whereas female participants had longer response times than male participants. Additionally, more mistakes were made when performing the memory task when the participant responded no to the trials compared to responding yes. Additionally, Appendix 14 shows an oveview of the mean response times for pressing ‘yes’ and ‘no’ for each stimulus, slightly fater resonse times were found when participants pressed ‘yes’ compared to pressing ‘no’.

Table 12. Pearson correlation matrix among selected variables.

		1	*2*	*3*	*4*	*5*	*6*	*7*
1	**Age**							
2	**Gender (0 = male, 1 = female)**	0.10						
3	**Self-assessed clarity rating**	-0.06	-0.16					
4	**Response behavior (0 = no, 1 = yes)**	0.05	0.01	-0.01				
5	**Response time (ms)**	0.08	**-0.21**	0.04	-0.14			
6	**Pupil diameter (mm)**	**-0.29**	0.01	0.02	0.03	-0.12		
7	**Memory task digits correct (0 = incorrect, 1 = correct)**	-0.15	-0.02	-0.05	**0.25**	-0.11	-0.09	
	Mean	23.48	0.69	7.79	0.49	1572	4.00	0.86
	Standard deviation	1.96	0.46	2.52	0.50	792	0.13	0.34

Note N = 103 for all variables.

4. DISCUSSION

We performed a photo-based computer experiment, combined with eye-tracking, to investigate the effect of eHMI message perspective and cognitive load on participants' perspective-taking, as inferred from their crossing intentions, when interacting with autonomous vehicles. Two eHMIs were designed to communicate that it was safe to cross and two to communicate that it was not safe to cross, without ambiguity. The ambiguous messages were designed in such a way that they can be interpreted in both ways; safe to cross or not safe to cross.

We hypothesized that egocentric messages would be more persuasive than allocentric messages, and that allocentric messages would be more persuasive than ambiguous messages. Additionally, we expected that the interpretation of egocentric messages would be less cognitively effortful than allocentric and ambiguous messages and, therefore, result in faster response times. Furthermore, we expected that when participants performed a concurrent memory load task, the clarity score would decrease, and response times would increase for all eHMI types though a more evident effect for allocentric messages. Lastly, we expected that pedestrians would take an egocentric perspective when interpreting ambiguous messages.

4.1. Main findings

The results show that the egocentric 'WALK'/DON'T WALK' were most persuasive compared to allocentric 'BRAKING/DRIVING' and the ambiguous 'GO/STOP', confirming our first hypothesis. Although, fastest response times were found for the egocentric messages when no memory task was present, no distinct effect between message perspective was found when a memory load was added. Also, eye-tracking data did not reveal any difference in pupil diameter between the egocentric and allocentric messages. The more ambiguous the message was regarded, the longer the time needed to interpret the message, suggesting that the time needed to interpret the meaning of the message depends more on the clarity rather than message perspective of the eHMI design; hence we could not confirm nor reject the second hypothesis (H2). Among the allocentric and ambiguous messages, no distinct effect was found; the allocentric 'DRIVING' and ambiguous 'GO', resulted in faster and more uniform crossing intentions compared to the allocentric 'BRAKING' and 'STOP'. Though a slightly (not significant) bigger pupil size was measured for the ambiguous messages at the moment of response, suggesting an increased perceived load compared to the egocentric and allocentric messages; therefore we could not confirm nor reject H4. Regarding the perception of pedestrians towards ambiguous eHMIs, we found that the majority used an egocentric perspective when making a crossing decision accepting H5. Contrary to our expectations, participants made faster crossing decisions while the crossing intentions were not affected when performing a concurrent memory task, hence rejecting H3 and H6.

4.2. Effect of message perspective

4.2.1 Clarity and crossing intentions

In accordance with some of the previous research (Ackermann et al., 2019; Bazilinskyy et al., 2019; De Clercq et al., 2019; Fridman et al., 2017; Qin, 2019), the egocentric 'WALK' and 'DON'T WALK' were the most favored eHMI designs as compared to the allocentric 'BRAKING' and 'DRIVING' and the ambiguous 'GO' and 'STOP', as evidenced by the higher self-reported clarity rating. The fact that egocentric messages were regarded as most clear might be due to the egocentric bias (as explained in the introduction), pedestrian or familiarity heuristics (Keysar et al., 2000; McDougall et al., 1999). Explicit and egocentric gestures (from the pedestrian's point of view) are already used in traffic situations to resolve ambiguity (e.g., using a hand gesture to give right of way, traffic signs using a green walking pedestrian or the text 'walk'); therefore the participants were already provided with a familiar type of message, while allocentric messages are often novel designs (McDougall et al., 1999). As expected, the ambiguous messages were perceived as most ambiguous, which is consistent with previous literature (Fridman et al., 2017). Some participants indicated, after completing the experiment, not understanding the perspective of the ambiguous messages; they were confused as to whether the messages were meant for them or provided information about the state of the vehicle.

In line with the self-assessed clarity rating, egocentric messages resulted in a better understanding, as evidenced by more uniform crossing intentions and compared to the allocentric and ambiguous messages. Another explanation for a higher mutual understanding of egocentric messages is that egocentric messages give the pedestrian an explicit instruction, leaving little room for interpretation (Ackermann et al., 2019; Fridman et al., 2017). However, it should be noted that among the egocentric and allocentric messages, the effect of message perspective was more distinct for the yielding vehicles compared to non-yielding vehicles. For non-yielding vehicles, the message 'DON'T WALK' only resulted in a higher clarity score compared to the allocentric 'DRIVING' during the baseline condition. Furthermore, the learning curves (as shown in Appendix 7), show that at the end of the experiment, the crossing intentions were even more uniform for 'DRIVING'. These findings suggest that explicitly showing non-yielding behavior was more important than the message perspective. For the yielding vehicles, the effect size of message perspective was large between the egocentric 'WALK' and the allocentric 'BRAKING'. The results showed that the ambiguous 'GO' even resulted in a higher clarity compared to the message 'BRAKING'. Even though the message 'BRAKING' was intended to be clear, 19% of the participants indicated not to cross, indicating participants still had difficulty understanding the meaning of the message. It remains unclear whether participants misinterpreted the message or whether the message was not explicit enough in this situation. The meaning of the message 'BRAKING' might be confusing, as participants may not know whether a yielding vehicle means that the vehicle is actually stopping. These results are in contrast with earlier findings, where the message 'BRAKING' was found to be clear (Deb et al., 2018). These contrary findings might be due to the difference in research method; in the present study, only images were used while in the study of Deb et al. (2018), vehicle dynamics were included as well. In the current experiment, the participant could not make use of other cues such as relative distance or vehicle speed to disambiguate the meaning of the message, which appears to be important factors for pedestrians to understand the behavior of the vehicle (Clamann et al., 2017). Since the allocentric 'BRAKING' and the ambiguous 'STOP' were found to be most unclear, we argue that both message perspective, as well as the explicitness of the message (e.g., no room for interpretation), are important to consider when designing an eHMI. These findings suggest that one might be used to receive egocentric messages from the driver when the vehicle is yielding, and think they are intended for them.

4.1.3. The perception of pedestrians towards ambiguous eHMIs

Research indicates that humans tend to rely on their own perspective and that switching from perspective costs cognitive processing time and effort (Keysar et al., 2000). Among the ambiguous messages, the majority of the participants interpreted the message egocentrically, pointing towards an egocentric bias or familiar heuristics, as mentioned before. Furthermore, a higher percentage interpreted the message 'GO' egocentrically compared to 'STOP', meaning that some participants intended to cross the street for both the messages. These results suggest that people tend to interpret an ambiguous message sooner as safe to cross. Research has shown that pedestrians tend to think that an AV will always give right of way. In a study by Dietrich et al. (2018), pedestrians significantly crossed earlier presented with any kind of information when encountering an AV. This might be an explanation for less uniform crossing intentions and thus a significantly higher percentage interpreting 'STOP' allocentrically compared to 'GO'. In the study of Fridman et al. (2017), 50% interpreted the message 'GO' egocentrically, and 50% interpreted the message 'stop' allocentrically.

4.1.2 Cognitive processes

In this study, we could not confirm whether perspective-taking is influenced by the use of different message perspectives. We expected that interacting with an explicit message perspective would be cognitively less effortful compared to ambiguous messages, with egocentric messages being easiest to understand; However, no distinct effect was measured. The results show no clear difference in cognitive processing time between the different message perspectives. When no memory task was present, egocentric messages resulted in the fastest response times, however when a memory task was present, about equal response times were measured for the egocentric messages and the allocentric 'DRIVING' and ambiguous 'GO', while only longer response times were measured for 'BRAKING' and 'STOP'. Although egocentric messages did result in (slightly) faster response times compared to allocentric messages (for the baseline condition), this does not indicate that it is due to switching from perspective. These results are in line with previous results, where no difference in response time was found between different message perspectives (Clamann et al., 2017; Deb et al., 2018). Our findings show that response times are highly correlated with the self-assessed clarity and the clarity score, indicating that the more ambiguous the message was regarded, the longer the time needed to interpret the message. According to the clarity scores, the message 'BRAKING' was more ambiguous than the egocentric messages and the allocentric message 'DRIVING', which was confirmed by the response times as well; that is, longer response times were found for the message 'BRAKING'. The same pattern was found for the ambiguous messages 'GO' and 'STOP', where the message 'GO' was perceived as less ambiguous than the message 'STOP' and resulted in shorter response times. In other words, the time needed to make a 'crossing decision' rather depended on the explicitness/clarity of the eHMI design than the perspective of the message itself. Hence, the more ambiguous the message was regarded, the longer the response times.

No significant differences were found for the different message perspectives regarding the pupil size. Though, for the ambiguous messages, the pupil size was slightly bigger compared to the egocentric and allocentric messages. These results indicate that the use of an ambiguous perspective was cognitive slightly more demanding than the use of an explicit perspective.

4.3. The effect of memory task

Contrary to our expectations, the crossing intentions of participants were hardly affected when performing a concurrent memory task. Furthermore, instead of longer response times, we even found that participants made faster crossing decisions when performing a memory task, with the most significant effect size for the message 'DRIVING' and 'GO'. Dual-process theories indicate that there are two different modes of processing – where one is fast, automatic, and heuristic-based, and one is slower, constrained by working memory (Kahneman & Frederick, 2002). Response time decreased when making a crossing decision when performing a memory task, suggesting that participants made use of an intuitive approach when being put under cognitive load. When a memory task was present, less working memory was available when performing the keypress task, and therefore participants had less ability to get distracted (Lavie & De Fockert, 2005). Hence, when performing a concurrent memory task during the experiment, participants had less ability to get distracted, resulting in faster response times. Additionally, at the moment of response, no difference in pupil diameter was measured for the different memory loads. These results suggest that participants were using efficient strategies to divide their attention since participants were able to maintain their performance (consistent crossing intentions) when a memory task was added. No change in mental effort was measured when making a crossing decision. Moreover, the fact that the interpretation of the meaning of the eHMI was not influenced when being put under cognitive load while participants were able to process the information faster is promising. When a pedestrian encounters a more complex traffic situation, and therefore the cognitive load increases, they still will be able to interpret the meaning of the eHMI correctly. Though, it should be noted that we used a memory task to increase the cognitive load, whereas, in real traffic, the cognitive load will also be increased due to visual load. However, the effect of memory task became less effective on the response times as a function of the trials; more specifically, a ceiling effect arose at the end of the experiment. The fact that a ceiling effect was achieved indicates that the memory task did not yield to increase the cognitive load during the whole experiment.

The response times of participants were most affected by a memory task for the allocentric 'DRIVING' and the ambiguous 'GO'. These findings suggest that perspective-taking did not occur when interpreting different message perspectives If perspective-taking were used to make crossing decision, it was expected that the interpretation of allocentric messages would have been more difficult under cognitive load due to the egocentric bias. For the egocentric messages, the decrease in response times was lower, which might be due to the fact that these messages are explicit, and therefore, without a memory task, people are already able to make a fast crossing decision. For the message 'DRIVING' and 'GO', it shows that participants used a more intuitive approach when performing a concurrent memory load task. For the messages that were perceived as most ambiguous (e.g., ‘STOP’ and ‘BRAKING’), the memory task had a low impact on the response times. Only when performing a high load memory task, the response times decreased slightly. These results indicate that participants were less able to use an intuitive approach for these messages, and a more controlled and effortful process was needed to disambiguate the meaning of the message.

Performance on the memory task shows that participants had more difficulty performing a low load versus a high load memory, as evidenced by a lower percentage correctly recalling the digits. As mentioned in the introduction, the pupil diameter can be used as an indication of cognitive load, whereas the pupil diameter increases with task difficulty (Hess & Polt, 1964). An increase in cognitive load was measured when the participants were presented with the digits; however, at the moment of responding, there was hardly any difference in pupil diameter compared to the baseline condition. When participants performed a memory task, the data showed that participants indeed perceived higher cognitive load compared to the baseline condition. Although we did find a significant difference in pupil diameter when participants performed a memory task, the pupil diameter did not differentiate at the onset of response. Because we found a difference in response times when participants performed a memory task, we can conclude that the memory load did affect the response behavior of the participants. However, in our experiment, the eye measurements stopped at the moment the participant had responded. Previous research found that the effect of a distractor in pupillary responses were later than the onset of the responses (Laeng, Ørbo, Holmlund, & Miozzo, 2011). This might also explain the fact that the pupil diameter was bigger at the beginning of each trial relative to the moment of response. The delay in pupillary response might be the reason for not finding differentiation in pupil diameter between the different eHMIs and memory load task at the moment of response. As can be seen in Figure 7, we also found that the pupillary response is slightly delayed when the digits were presented. For example, at t=1, the first digit was presented; around 600ms later, a pupillary response was found, e.g., the pupil diameter increased.

Summarizing, egocentric messages were regarded as most clear as evidenced by a higher perceived safety and more uniform crossing intentions compared to allocentric and ambiguous messages. Furthermore, when the message perspective was ambiguous, participants used an egocentric interpertation, pointing towards an egocentric bias. Altough, increased cognitive load was measured for memory task (as demonstrated by increased pupil diameter when the digits were presented), no difference in pupil diameter was found at the moment of response. These findings suggest that when participants were put under cognitive load, an efficient strategy was used to make a crossing decision as evidenced by the consistent crossing intentions but faster response times.

5. CONCLUSION

The goal of this study was to investigate whether message perspective and cognitive load influence perspective-taking when pedestrians make a crossing decision based on information presented by an eHMI. When the message perspective is ambiguous, the message ‘GO’ encouraged crossing whereas the message ‘STOP’ inhibited crossing, pointing towards and egocentric perspective taken by the participants. In addition, a higher percentage interpreted the message ‘GO’ egocentrically compared to ‘STOP’, meaning that some participants intended to cross the street for both the message ‘GO’ and ‘STOP’, suggesting that pedestrians tend to think that an AV will always give right of way. Furthermore, regarding the use of message perspective, egocentric messages were regarded as clearest as evidenced by more uniform crossing intentions, higher perceived safety and faster decision times. Furthermore, the response times were highly correlated with the objective and subjective clarity, indicating that response times are and index for perceived the clarity of the eHMI. When a memory task was present, participants had less memory capacity available to get distracted and therefore made use of a more efficient strategy, resulting in faster crossing decision. participants made faster crossing decisions while the crossing intentions were not affected when performing a concurrent memory task, suggesting that participants used an efficient strategy to make a crossing decision when being put under cognitive load. Concluding, it can be deduced that pedestrians initially take an egocentric perspective if the eHMI message is ambiguous, though this egocentric bias can be overcome by using explicitly an egocentric or allocentric eHMI message perspective. In addition, we conclude that participants perform better (more uniform crossing decisions, faster responses) when the eHMI’s message perspective is egocentric rather than allocentric.

6. Limitations and Recommendations for future research

For this research, the six eHMI designs have been compared in a within-subjects design using a total of 103 participants. The participant viewed images of vehicles equipped with an eHMI and was asked to press shift (L-shift = no, R-shift = yes) to indicate whether they thought they could cross, while their eye movements were measured with an eye tracker. As mentioned in the result section, data from 62 participants, which amounts to approximately one third of the total number of participants, were excluded. Although we tried to compile the instructions for the experiment as clearly as possible, the experiment appeared to be unclear for a large number of participants. One of the main reasons for misunderstanding during the experiment was due to not reading the instructions carefully enough. After the experiment, a number of participants indicated not to have seen any images of the vehicle, after which it appeared that the participant had pressed 'yes' (R-shift) when the statement 'I can cross' was shown. Furthermore, some participants indicated after the experiment not understanding the experiment at all or using the wrong keys. Apparently, the experiment was designed in such a way, that after pressing any key at any moment, the participant was already redirected to the next trial. Ideally, the participant would only proceed to the next trial when the participant used the appropriate keys. Furthermore, the participants should not have been proceeded to the next trial when pressing a key before the onset of the image. However, only during the experiment, we found out that the settings of the experiment were not set correctly. In addition, some participants indicated not to understand the task of the experiment during the first couple of trials or asked for additional explanation while they had already started the experiment. Furthremore, over a sequence of the 18 trials, participants showed an increase in clarity score, indicating that there was some learning behavior regarding the experiment (see Appendix 7). Several participants indicated that they did not fully understand the task in the first few trials, explaining the increase of correct answers. As a result, the clarity score and response times may not match the actual clarity of the eHMI.

Another limitation of this study is that we only used textual messages to investigate the effect message perspective. Although textual displays are often found to be most evident, some researchers are skeptical about whether the use of textual displays will be understandable for different cultures. First of all, with the use of textual displays, it should be considered that not everyone speaks English and should, therefore, be universal, which is a major drawback from using text. Therefore, symbolism and color-coding have been argued to be more effective as they overcomes the language barrier (Dietrich et al., 2018). Also, in traffic, symbolic displays have a higher understandability and are easier to interpret from a distance than textual displays (Kline, Braun, Peterson, & Silver, 1993). Additionally, although the results show that textual displays were generally regarded as egocentric when there is clear message perspective available, this does not mean that all eHMIs will be interpreted egocentrically. A textual display might be interpreted from heuristic point of view (familiar with receiving textual information from traffic signs), while LED lights are more often associated with a sensor on the vehicle and thus associating this feature with the intention of the vehicle (Bazilinskyy et al., 2019). Thus, the perspective taken when interpreting the intention of the vehicle might also depend on the type of message coding. To further research the effect of perspective-taking in AV-P interaction, different types of message coding should be considered to get to obtain a deeper understanding. Then again, for this research, we aimed to shed light on the effect of message perspective with regards to perspective-taking and therefore used textual messages and focused less on whether textual messages should be used as an external device for autonomous vehicles.

Another limitation of our study is the use of a memory load to increase the cognitive load and thus distract them when making a crossing decision. When researching the impact of eHMIs on crossing behavior of pedestrians, the pressure or incentive accuracy experienced by the participant highly depends on the complexity of the traffic situation (the number of road users) as well as how the experiment is conducted. In order to make the traffic situation more realistic from a cognitive perspective, we added a memory load task to increase the cognitive load experienced during the trials. We can conclude that in our experiment, instead of distracting the participant, we actually accomplished that participant had less memory capacity available to get distracted. Furthermore, the memory task did not yield to increase the cognitive load during the entire experiment.

Another limitation is that we performed a computer-based experiment using photos. Whether pedestrians will indeed interpret the message ‘GO’ and ‘STOP’ egocentrically when the intention of the message is allocentric remains unclear. First of all, we only used photos during the experiment, limiting the participant to make use of other cues such as vehicle dynamics (which was also the case for the message ‘BRAKING’). Furthermore, participants needed to imagine whether they would cross or not from a static environment (behind the computer). Since participants are not actually at risk of making an incorrect crossing decision, which might result in a response bias (e.g., not matching actual behavior when encountering a self-driving vehicle in real traffic). Moreover, participants might have had less motivation to accurately make a correct decision when interpreting the ambiguous messages as they would have been in a more realistic environment, which also might have led to a more egocentrically interpretation (as explained in the introduction regarding the egocentric bias, Keysar, 2007).

This thesis attempts to get a deeper understanding whether pedestrians make use of perspective-taking when interpreting the intention of AVs equipped with eHMIs, as well as the effect of message perspective on perspective taking. Further research in dynamic environments and more complex situations are required before conclusions can be drawn about whether pedestrians make use of perspective taking as well as the effect of message perspective.

7. REFERENCES

Ackermann, C., Beggiato, M., Schubert, S., & Krems, J. F. (2019). An experimental study to investigate design and assessment criteria : What is important for communication between pedestrians and automated vehicles ? Applied ErGOnomics, 75(March 2018), 272–282. https://doi.org/10.1016/j.aperGO.2018.11.002

Avtanski, A. (2020). LCD display generator. Retrieved from http://avtanski.net/projects/lcd

Bazilinskyy, P., Dodou, D., & De Winter, J. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, *67*, 175-194. https://doi.org/10.1016/j.trf.2019.10.013

Beggiato, M., Witzlack, C., & Springer, S. (2018). The Right Moment for BRAKING as Informal Communication Signal Between Automated Vehicles and Pedestrians in Crossing Situations. https://doi.org/10.1007/978-3-319-60441-1

Bockle, A. P. (2017). Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians ' Experience SAV2P – Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians ' Experience, (October). https://doi.org/10.1145/3131726.3131765

Chang, C., Toda, K., Sakamoto, D., & Igarashi, T. (2017). Eyes on a Car : an Interface Design for Communication between an Autonomous Car and a Pedestrian, (Figure 1), 65–73.

Clamann, M., Aubert, M., & Cummings, M. L. (2017). Evaluation of Vehicle-to-Pedestrian Communication Displays for Autonomous Vehicles, 1–13.

Cousineau, D. (2005) Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method, Tutorials in Quantitative Methods for Psychology, 1(1), 42-45. https://doi.org/10.20982/tqmp.01.1.p042

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. Frontiers in Human Neuroscience, 7, 596. doi:10.3389/fnhum.2013.00596

Davis, M. H., Conklin, L., Smith, A., & Luce, C. (1996). Effect of perspective taking on the cognitive representation of persons: A merging of self and other. *Journal of Personality and Social psychology*, 70(4), 713-726. https://doi.org/10.1037/0022-3514.70.4.713

De Clercq, K., Dietrich, A., Núñez Velasco, J. P., De Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: effects on pedestrian crossing decisions . *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(8), 1353-1370. https://doi.org/10.1177/0018720819836343

Deb, S., Strawderman, L. J., & Carruth, D. W. (2018). Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59*, 135-149. https://doi.org/10.1016/j.trf.2018.08.016

Dey, D., & Terken, J. (2016). Pedestrian Interaction with Vehicles : Roles of Explicit and Implicit Communication, 109–113.

Dietrich, A., Willrodt, J., Wagner, K., & Bengler, K. (2018). Projection-Based external Human Machine Interfaces Enabling Interaction between Automated Vehicles and Pedestrians, (September).

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327-339. https://doi.org/10.1037/0022-3514.87.3.327

Färber, B. (2016). Communication and communication problems between autonomous vehicles and human drivers. *In Autonomous Driving* (pp. 125-144). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-48847-8_7

Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. The Quarterly Journal of Experimental Psychology, 70(8), 1646-1660.

Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., & Campione, J. C. (1986). Development of knowledge about the appearance-reality distinction. Monographs of the society for research in child development, i-87.

Fridman, L., Mehler, B., Xia, L., Yang, Y., Yvonne, L., & Bryan, F. (2017). To WALK or Not to WALK : Crowdsourced Assessment of External Vehicle-to-Pedestrian Displays

Habibovic, A., & Davidsson, J. (2012). Causation mechanisms in car-to-vulnerable road user crashes: Implications for active safety systems. Accident Analysis & Prevention, 49, 493-500.

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., … Wood, S. (2018). Communicating Intent of Automated Vehicles to Pedestrians, 9(August). https://doi.org/10.3389/fpsyg.2018.01336

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity*, *43*(1), 59-74.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. Science, 143(3611), 1190-1192.

Hudson, C. R., Deb, S., Carruth, D. W., McGinley, J., & Frey, D. (2018, July). Pedestrian Perception of Autonomous Vehicles with External Interacting Features. In International Conference on Applied Human Factors and ErGOnomics (pp. 33-39). Springer, Cham.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. Science, 154(3756), 1583-1585.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. Heuristics and biases: The psychology of intuitive judgment, 49, 81.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. Psychological Science, 11(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they 'should'. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 150–166). New York, NY, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511808098.010

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. Intercultural Pragmatics, 4, 71–84. https://doi.org/10.1515/IP.2007.004

Kline, P. B., Braun, C. C., Peterson, N., & Silver, N. C. (1993, October). The impact of color on warnings research. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 37, No. 14, pp. 940-944). Sage CA: Los Angeles, CA: SAGE Publications.

Kooijman, L., de Winter, J., (2019). The Detection of Pedestrian Crossing Behaviour. http://resolver.tudelft.nl/uuid:9cd6529f-039e-44ec-9451-72c18174e3e0

Lagstrom, T., & Lundgren, V. M. (2015). AVIP-Autonomous vehicles interaction with pedestrians. Master of Science Thesis, Chalmers University of Technology.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary stroop effects. Cognitive processing, 12(1), 13-21.

Lavie, N., & De Fockert, J. (2005). The role of working memory in attentional capture. Psychonomic bulletin & review, 12(4), 669-674.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. Journal of Experimental Social Psychology, 46(3), 551-556.

Lundgren, V. M., Habibovic, A., Andersson, J., Lagström, T., Nilsson, M., Sirkka, A., ... & Saluäär, D. (2017). Will there be new communication needs when introducing automated vehicles to the urban context?. In Advances in human aspects of transportation (pp. 485-497). Springer, Cham.

Matthews, M., & Chowdhary, G. V. (2017). Intent Communication between Autonomous Vehicles and Pedestrians.

McDougall, S. J., Curry, M. B., & De Bruijn, O. (1999). Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, *31*, 487–519. https://doi.org/10.3758/BF03200730

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Nissan. (2015). IDS Concept. Retrieved from https://www.nissan.co.uk/experience-nissan/concept-cars/ids-concept.html

Palmeiro, A. R., van der Kint, S., Vissers, L., Farah, H., de Winter, J. C., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. *Transportation research part F: traffic psychology and behaviour*, *58*, 1005-1020.

Petzoldt, T., Schleinitz, K., & Banse, R. (2018). Potential safety effects of a frontal brake light for motor vehicles, 449–453. https://doi.org/10.1049/iet-its.2017.0321

Qin, Z. (2019). *Design and Assessment of External Displays on Autonomous Vehicles for Pedestrian Safety* (Master's thesis, University of Waterloo).

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230-236.

Reiff, A., De Winter, J., & Eisma, Y., (2020). Perspective-taking in autonomous vehicle-pedestrian interaction – The effect of message perspective: A literature survey. ??

Roxβnagel, C. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, *30*(3), 429-445.

SAE International. 2014. Taxonomy and Definitions for Terms Related to On?Road Motor Vehicle Automated Driving Systems.

Santiesteban, I., Catmur, C., Hopkins, S., Bird, G., & Heyes, C.M., (2014). Avatars and arrows: Implicit mentalizing or domain general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*, 929-937.

Stanciu, S. C., Eby, D. W., Molnar, L. J., Louis, R. M. S., Zanier, N., & Kostyniuk, L. P. (2018). Pedestrians/Bicyclists and Autonomous Vehicles: How Will They Communicate? https://doi.org/10.1177/0361198118777091

Strayer, D. L., & Fisher, D. L. (2016). SPIDER: A framework for understanding driver distraction. *Human factors*, *58*(1), 5-12.

Sucha, M., Dostal, D., & Risser, R. (2017). Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention*, *102*, 41-50.

Todd, A. R., Cameron, C. D., & Simpson, A. J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, *159*, 97-101.

Turnwald, A., Althoff, D., Wollherr, D., and Buss, M. (2016). Understanding human avoidance behavior: interaction-aware decision making based on game theory. Int. J. Soc. Robot. 8, 331–351. doi: 10.1007/s12369-016-0342-2

Vukovic, N., & Shtyrov, Y. (2017). Cortical networks for reference-frame processing are shared by language and spatial navigation systems. *NeuroImage*, *161*(August), 120–133. https://doi.org/10.1016/j.neuroimage.2017.08.041

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019, July). Crossing the Street Across the Globe: A Study on the Effects of eHMI on Pedestrians in the US, Germany and China. In *International Conference on Human-Computer Interaction* (pp. 515-530). Springer, Cham.

Zhang, J., Vinkhuyzen, E., & Cefkin, M. (2018). Evaluation of an Autonomous Vehicle External Communication System Concept : A Survey Study, (1). https://doi.org/10.1007/978-3-319-60441-1

Zimmermann, R., & Wettach, R. (2017, September). First step into visceral interaction with autonomous vehicles. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 58-64). ACM.

8. APPENDICES

ACKNOLWEDGEMENTS 4

APPENDIX 1: INFORMED CONSENT 46

APPENDIX 2: STEPS AND INSTRUCTIONS 48

APPENDIX 3: OVERVIEW OF EXPERIMENT; ONE OF THE 18 STIMULI USED IN THE EXPERIMENT: EHMI CONCEPT STOP, HIGH LOAD MEMORY TASK 50

APPENDIX 4: OVERVIEW OF ALL STIMULI 51

APPENDIX 5: QUESTIONNAIRE 52

APPENDIX 6 UNPROCESSED DATA RESPONSE BEHAVIOR 55

APPENDIX 7 LEARNING BEHAVIOR 56

APPENDIX 8 NUMBER OF TIMES APPEARANCE OF STIMULI FOR EACH TRIAL 58

APPENDIX 9 RESPONSE TIME AS A FUNCTION OF TRIAL NUMBER FOR EACH STIMULUS 59

APPENDIX 10 PERCENTAGE PRESSING 'YES' AS A FUNCTION OF TRIAL NUMBER FOR EACH STIMULUS 60

APPENDIX 11 MEAN DIFFERENCE IN RESPONSE TIMES FOR MEMORY LOAD 61

APPENDIX 12 MEAN PUPIL DIAMETER AS A FUNCTION OF TIME FOR MESSAGE PERSPECTIVE AND EHMI 62

APPPENDIX 13 OVERVIEW OF POST-HOC COMPARISONS OF MEMORY TASK FOR ALL TIME BINS 64

APPENDIX 14: MEAN RESPONSE TIME (MS) FOR PRESSING 'YES' AND 'NO' FOR EACH STIMULUS 65

APPENDIX 15: INDIVIDUAL RESULTS FOR RESPONSE TIME AND CROSSING INTENTIONS 66

APPENDIX 1: INFORMED CONSENT

Consent form for participants

Eye movements while performing visual/spatial tasks

Researchers:
ir. Yke Bauke Eisma: y.b.eisma@tudelft.nl
ir. Lars Kooijman: l.kooijman-1@tudelft.nl
Anne Reiff: A.L.Reiff@student.tudelft.nl
Dr. Dimitra Dodou: d.dodou@tudelft.nl
Dr. ir. Joost de Winter: j.c.f.dewinter@tudelft.nl

Location of the experiment
Room 34 A-0-811 (above lecture room C), see Figure 1.
Faculty of Mechanical, Maritime and Materials Engineering Delft University of Technology
Mekelweg 2, 2628 CD Delft

Figure 1. *Location of the experiment (door at the right side, then up the stairs to the right)*

Note: The measurement equipment functions better with contact lenses than glasses. If possible, please wear contact lenses instead of glasses.

Introduction: Please read this consent form carefully before you decide to participate. This document describes the purpose, procedures, and potential risks/discomforts. Your signature is required for participation.

Purpose of the study: This study has three aims. The first aim is to investigate eye movements and pedestrian crossing decisions while viewing text messages presented on automated vehicles. The second aim is to investigate what eye movements are used while performing a perceptual speed task and to what extent such movements are associated with the responses of the individual. The third aim is to investigate what visual strategies are used while solving challenging visual-spatial problems.

Duration: Your participation in this experiment will last approximately 30 minutes.

Figure 2. *Experimental setup with head support and eye tracker*

Procedures and instructions

Before the experiment starts: You will be asked to rest your head on the support (see Figure 2) and look at specific places on the screen so that we can calibrate the eye-tracking equipment.

During the experiment: First, you will be asked to look at photos of automated vehicles and indicate whether you can cross or not. In some cases, you will be asked to conduct a memory task at the same time. Next, you will be asked to perform a perceptual speed task. Finally, you will be asked to solve a number of visual-spatial problems.

After the experiment: You will be asked to complete a short questionnaire about your gender, age, and glasses/contacts.

Risks and discomforts: There are no known risks for you in this study. Some minor eyestrain or discomfort may arise from the monitoring task. If at any point you begin to feel uneasy for any reason, please do not hesitate to inform the experimenter so that you take a break to counteract any such symptoms.

Anonymity: All data collected in this study will be stored in an anonymous manner. You will not be personally identifiable in any future publications based on this work and in any data files that may be stored in an online repository or shared with other researchers.

Right to refuse or withdraw: Your participation in this study is entirely voluntary. You have the right to refuse or withdraw from this experiment at any time, without any negative consequences, and without needing to provide any explanation.

Questions: For any questions, you can contact one of the researchers at the email addresses above.

I have read and understood the information provided above. I give permission to store and use of collected data for the purposes of this study described above. The results of the study will not be made available in a way that could reveal the identity of individuals. I voluntarily agree to participate in this study.

Name:

…………………………………………………………………………………………………..

Signature:

…………………………………………………………………………………………………..

Date:

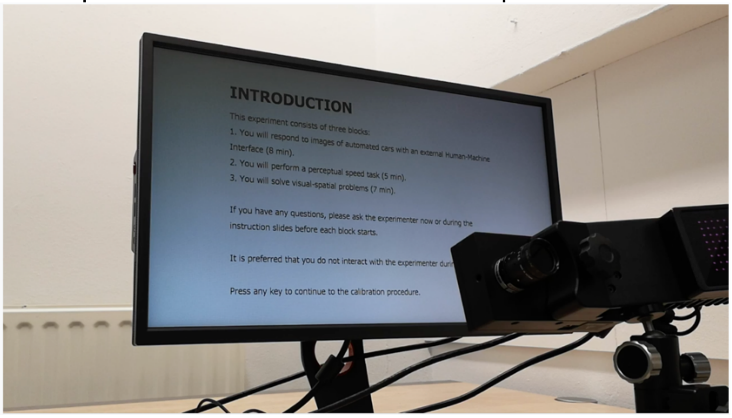
……… / …….. / ………

APPENDIX 2: STEPS AND INSTRUCTIONS

The experiment consists of three blocks, of which only the first block of the experiment is part of this study.

1. Arrival participant
2. Briefing experiment using informed consent
3. Signing informed consent

4. Participant takes place behind the computer
5. Participant reads introduction to the experiment

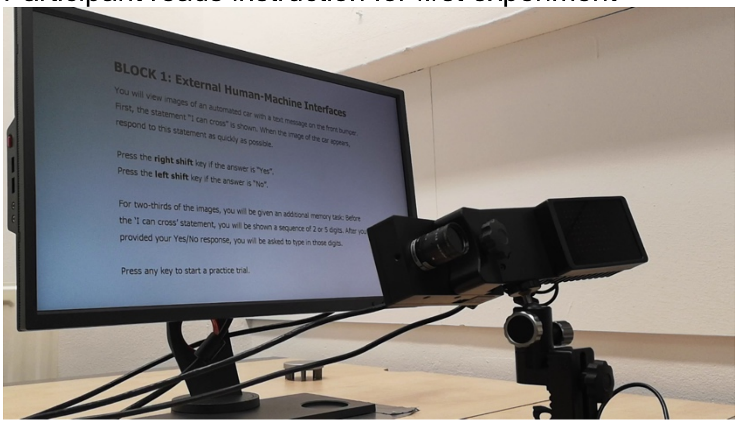


6. Calibration

7. Reminder to participant before starting the experiment regarding
   a. carefully read the instructions for each experiment
   b. keep head in head support during the whole experiment

Start block 1

8. Participant reads instruction for first experiment



9. Start familiarization trial
10. Start experimental trial

11. After 18 trials, experiment ends with questionnaire

Start block 2 and 3 (not part of this study)

12. Start second block (perceptual speed task)
13. Participant reads instructions, familiarization trial, starts experiment
14. After completing second experiment, start third block (visual-spatial problems)
15. Participant reads instructions, familiarization trial, starts experiment

After finishing all three blocks

16. Demographic questionnaire by participant
17. Answer questions participant
18. Departure experiment

Appendix 3: Overview of experiment; one of the 18 stimuli used in the experiment: eHMI concept STOP, high load memory task

1

2

3

4

5

6

7

8

9

10

APPENDIX 4: OVERVIEW OF ALL STIMULI

PERSPECTIVE	Memory load	eHMI
Egocentric	Baseline: - Low load: 16 High load 95246	
	Baseline: - Low load: 47 High load: 63908	
Allocentric	Baseline: - Low load: 36 High load: 74524	
	Baseline: - Low load: 17 High load: 74524	
Ambiguous	Baseline - Low load: 97 High load: 52672	
	Baseline - Low load: 02 High load: 29312	

Note: The digits indicate which digit the participant had to remember during for each eHMI. As mentioned in the introduction, the digits were fixed; meaning that each participant performed the same memory load.

APPENDIX 5: QUESTIONNAIRE

In the following slides, you are asked questions about the external Human-Machine Interfaces that you have seen.

Press any key to continue to the questions

Please indicate on a scale from 0 to 10 (0 = completely disagree, 10 = completely agree)

The message on the vehicle is clear

…

Press enter to continue

Please indicate on a scale from 0 to 10 (0 = completely disagree, 10 = completely agree)

The message on the vehicle is clear

…

Press enter to continue

Please indicate on a scale from 0 to 10 (0 = completely disagree, 10 = completely agree)

The message on the vehicle is clear

…

Press enter to continue

Please indicate on a scale from 0 to 10 (0 = completely disagree, 10 = completely agree)

The message on the vehicle is clear

…

Press enter to continue

Please indicate on a scale from 0 to 10 (0 = completely disagree, 10 = completely agree)

The message on the vehicle is clear

…

Press enter to continue

Please indicate on a scale from 0 to 10 (0 = completely disagree, 10 = completely agree)

The message on the vehicle is clear

…

Press enter to continue

APPENDIX 6 UNPROCESSED DATA RESPONSE BEHAVIOR

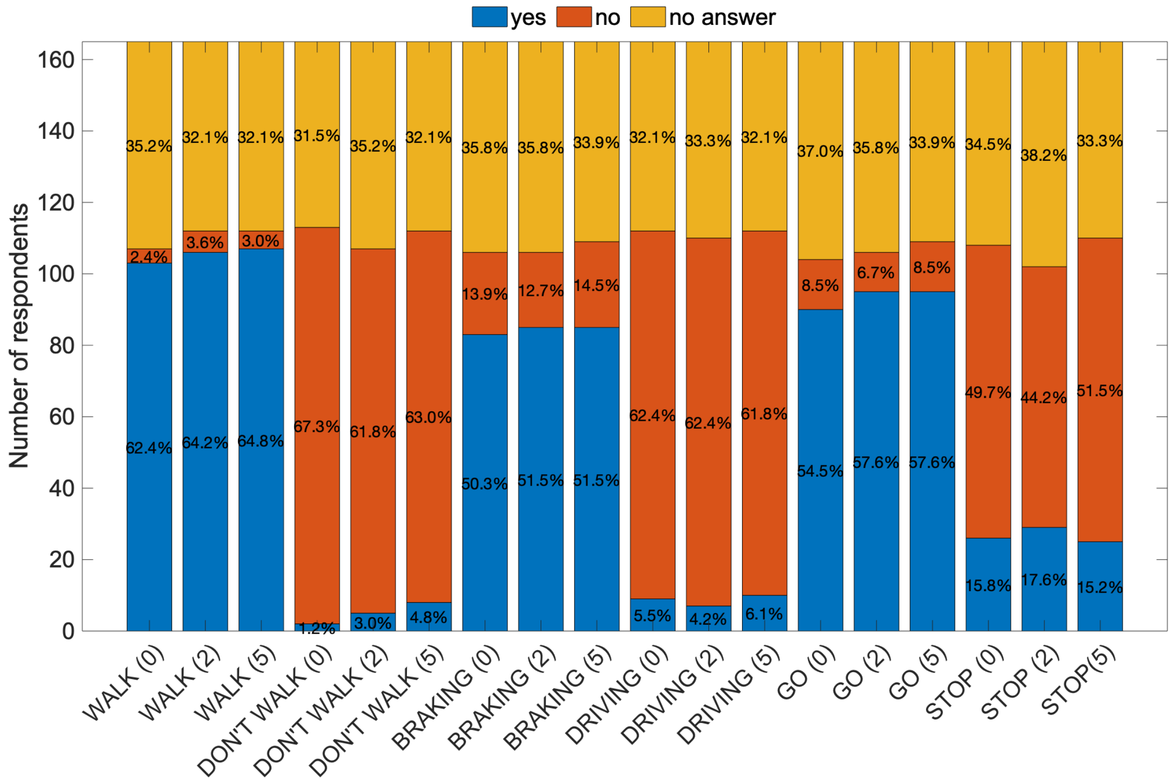


Figure 9. Distribution of answers to the statement 'I can cross' for each trial before processing the data. The number between parentheses indicates the number of digits during the memory task. 0 stands for no memory task, 2 stands for the 2-digit memory task and 5 stands for the 5-digit memory task. The percentage is calculated based on the number of respondents for each trial.

APPENDIX 7 LEARNING BEHAVIOR

The effect of message perspective

Participants encountered 18 trials, with each eHMI three times under different memory load during the experiment. Figure 10 provides illustrative learning curves, which were fit using the following function: y = a*exp(b*x)+c. As for response behavior and response times, a learning effect can be identified. An increasing overall clarity score was found for each eHMI except for the message 'DON'T WALK', indicating that for the other messages the number of participants answered 'yes' for the yielding vehicles and 'no' for the non-yielding vehicles increased as the trials progressed (Figure 10). As mentioned in the result section, some participants indicated not to understand the experiment in the first couple of trials or asked for clarification after already starting the experiment, which is one of the reasons for an increase in clarity. However, the figure does show some different response behavior regarding the eHMIs, where a longer learning effect can be identified for the ambiguous messages compared to the allocentric and egocentric messages.

Furthermore, for the egocentric messages, the shortest learning effect is found, whereas for the message 'DON'T WALK', no learning effect at all was found. However, not too much value should be attached to these results. In Appendix 8 an overview of the number of times a particular stimulus has appeared per trial is presented. Looking at the message 'DON'T WALK', it can be seen that the message was only presented four times for the first trial (baseline: once, low load: zero times, high load: three times). However, the clarity score for the message remained stable after the first trial, indicating that the message 'DON'T WALK' did not required any learning. In other words, the message 'DON'T WALK' was perceived as most intuitive, followed by the message 'WALK', whereas the ambiguous messages appeared to be least intuitive.

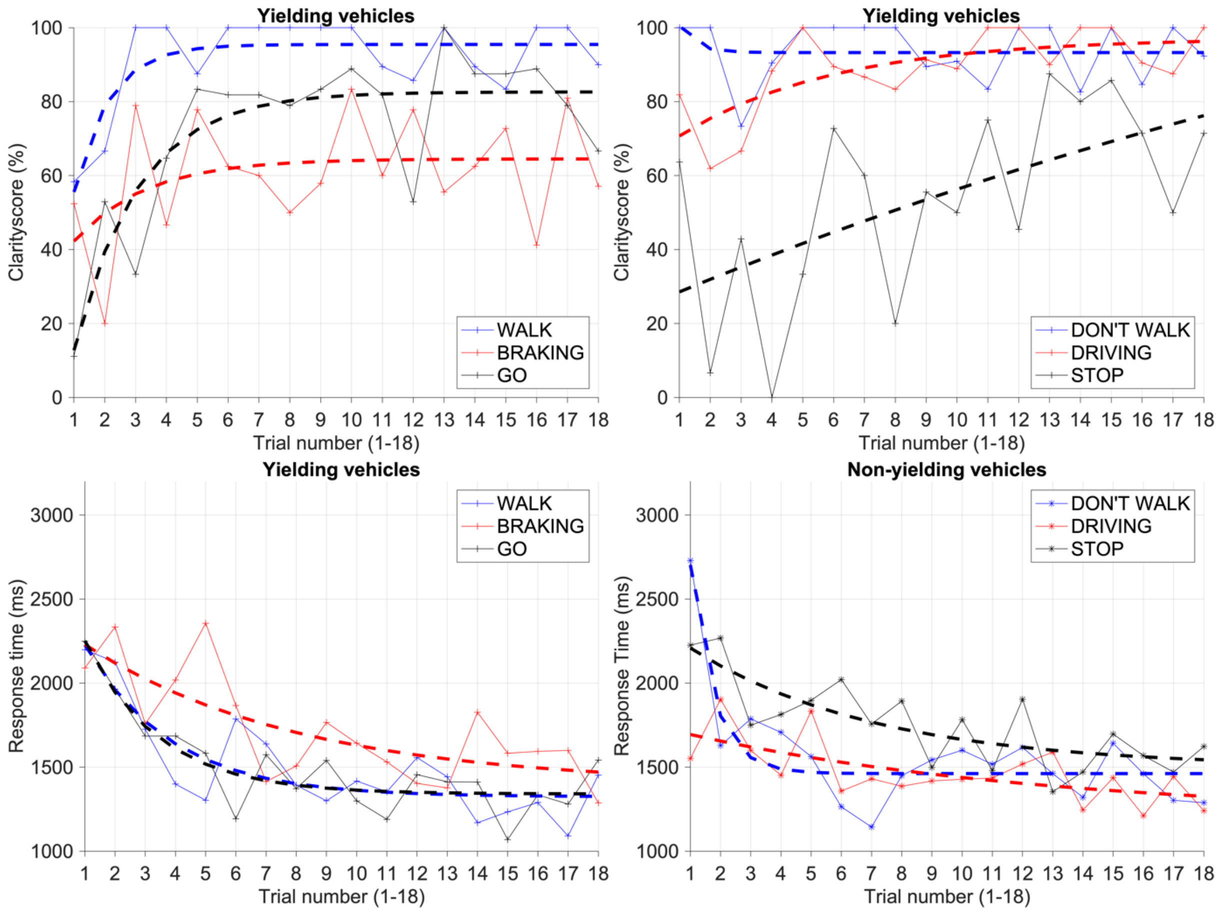


Figure 10. Learning behavior for all eHMIs for clarity score (e.g., response behavior) and response time. The stars are the mean for each trial for response time and response behavior.

As for the response time, also a learning effect can be identified; faster response times were found as a function of the trials (Figure 9). For the yielding vehicles, it can be seen that the response time at the beginning of the experiment is the same for each message. However, faster response times were found for the message "WALK" and "GO" than for the message "BRAKING" at the end of the experiment, indicating tthat participants still had more difficulty interpreting the allocentric ‘BRAKING’ at the end of the experiement. Forn non-yielding vehicles, relatively long response time for the message ‘DON’T WALK’ was found during the first trial; as a result, a substantial learning effect is seen. However, as explained in the previous section, the message only appeared four times in the first trial, so the results are based on a relatively small sample. If the first trial is not included, the curve of "DON'T WALK" will roughly coincide with the curve of the message "DRIVING". Having said that, it can be seen that the response time at the beginning of the trial is low for the message ‘DON’T WALK’ and ‘DRIVING’, and there is a slight decrease in response time across the trials, which means that there was little learning effect. For the message "STOP" a longer learning curve can be seen, which corresponds to the curve of the message "BRAKING". In other words, for yielding vehicles, no learning effect was found for the message "DRIVING" and "DON'T WALK," and fast learning behavior was found for the yielding vehicles "WALK" and "GO". The message "BRAKING" and "STOP", the messages that were perceived as most ambiguous, also required more learning.

We argue that the familiarization of the experiment mostly causes the learning effect regarding the response behavior for egocentric and allocentric messages. In contrast, for the ambiguous messages, the learning effect is also caused by the ambiguity of the message. In other words, egocentric and allocentric messages required less learning than ambiguous messages, indicating that explicit use of message perspective has a positive impact on learning behavior. As for the response time, we argue that the learning effect is caused by the familiarization as well as the perceived clarity of the message.

The effect of memory load

When looking at the clarity scores, a learning effect is identified for all memory load conditions. However, no substantial difference has been for pedestrians crossing intentions as a function of the trials between the different memory load conditions. fFor response times, also a learning effect is identified for all memory load conditions. Interestingly, when looking at the response times, it can be seen that the memory task became less effective as a function of time (e.g., ceiling effect). At the end of the trial, response times were equal for all memory task conditions, indicating that the independent variable memory load did not have an effect on the dependent variable response time anymore.

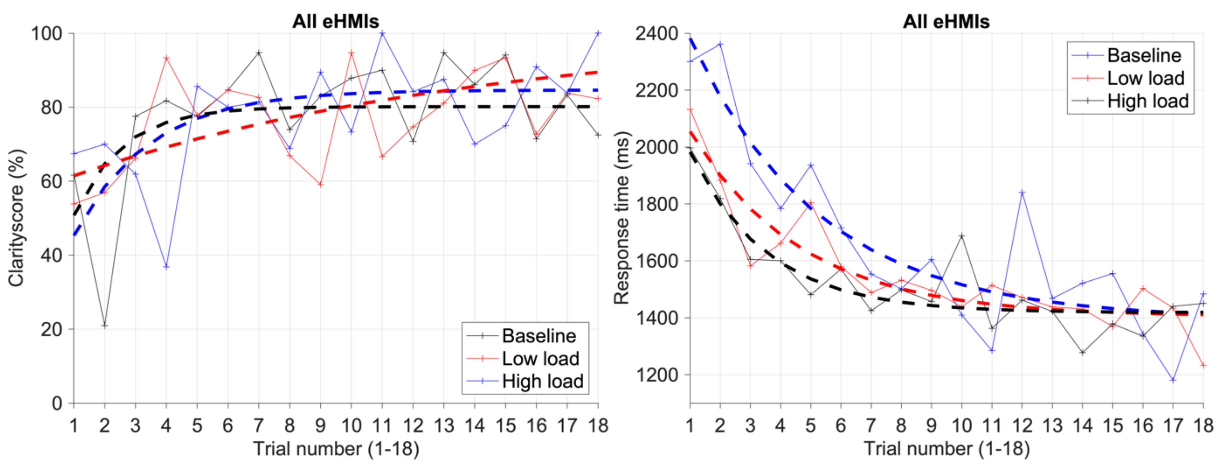


Figure 11. Learning behavior for memory task for clarity score (e.g., response behavior) and response time. The stars are the mean for each trial for response time and response behavior.

APPENDIX 8 APPEARANCE OF EACH STIMULI FOR EACH TRIAL

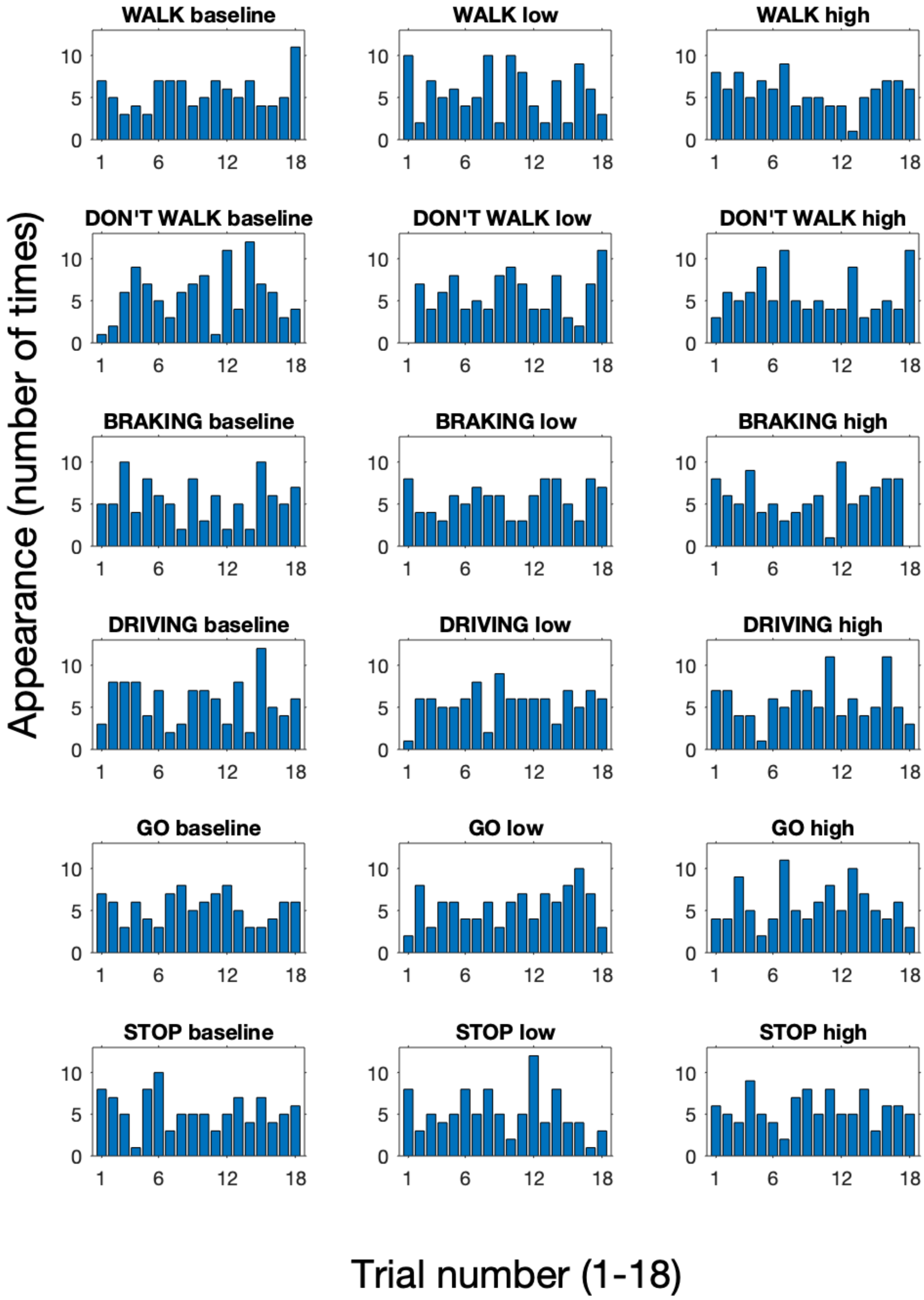


Figure 12. An overview of the appearance of the 18 stimuli during the experiment. The number of times each stimuli was shown for each trial during the experiment. For example, the figure shows that the stimulus 'DON'T WALK' with a low memory load, was not presented during the first trial during the whole experiment.

APPENDIX 9 RESPONSE TIME AS A FUNCTION OF TRIAL NUMBER FOR EACH STIMULUS

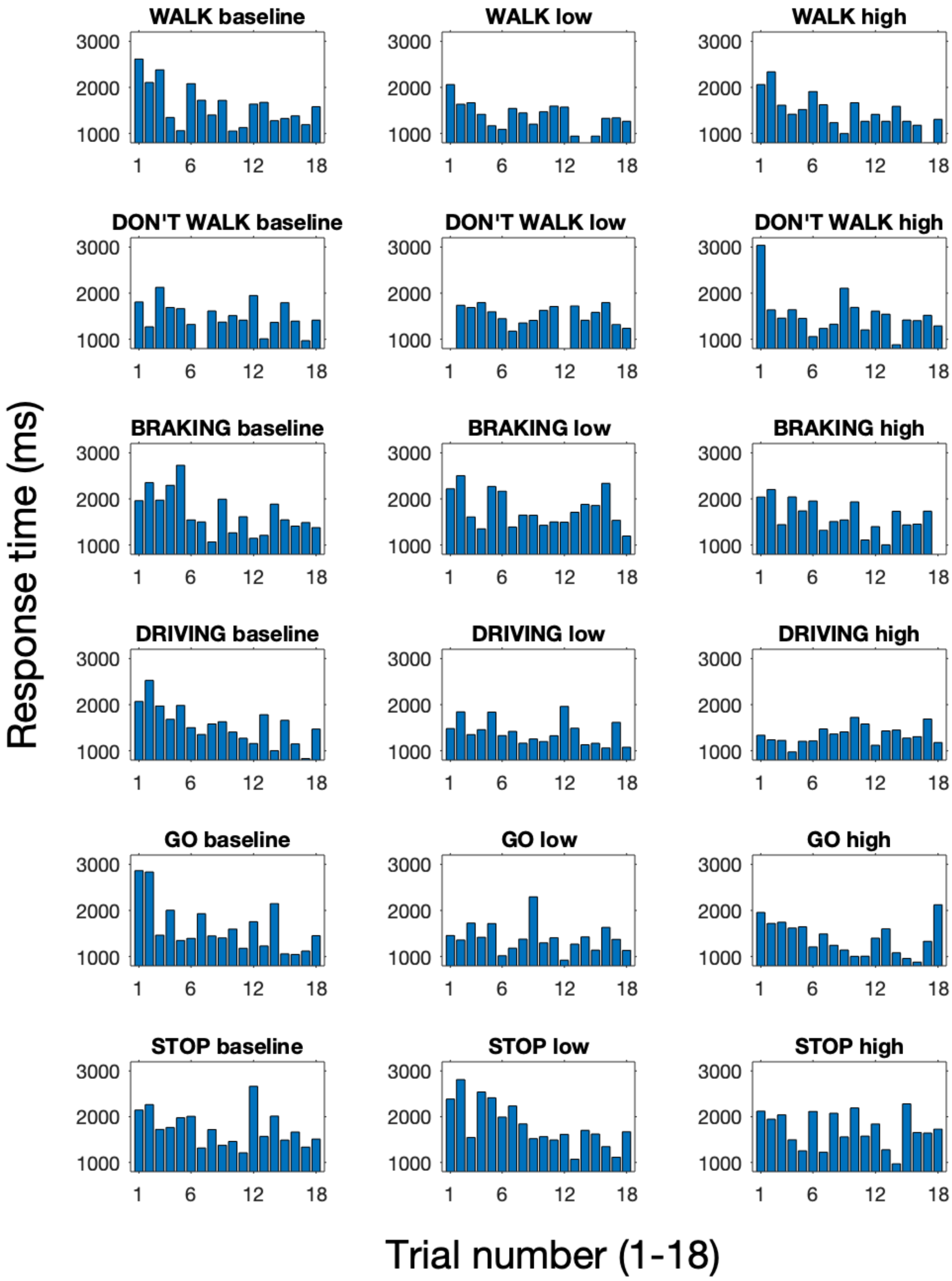


Figure 13. An overview of the mean response times for each stimuli for each trial during the experiment. The column shows the memory task (baseline, low load and high load) and the rows show the six external Human-Machine Interfaces (eHMIs). For example, the figure shows that the response time for the stimulus 'DON'T WALK' with a high memory load was about one second higher compared to the other trials.

APPENDIX 10 PERCENTAGE PRESSING 'YES' AS A FUNCTION OF TRIAL NUMBER FOR EACH STIMULUS

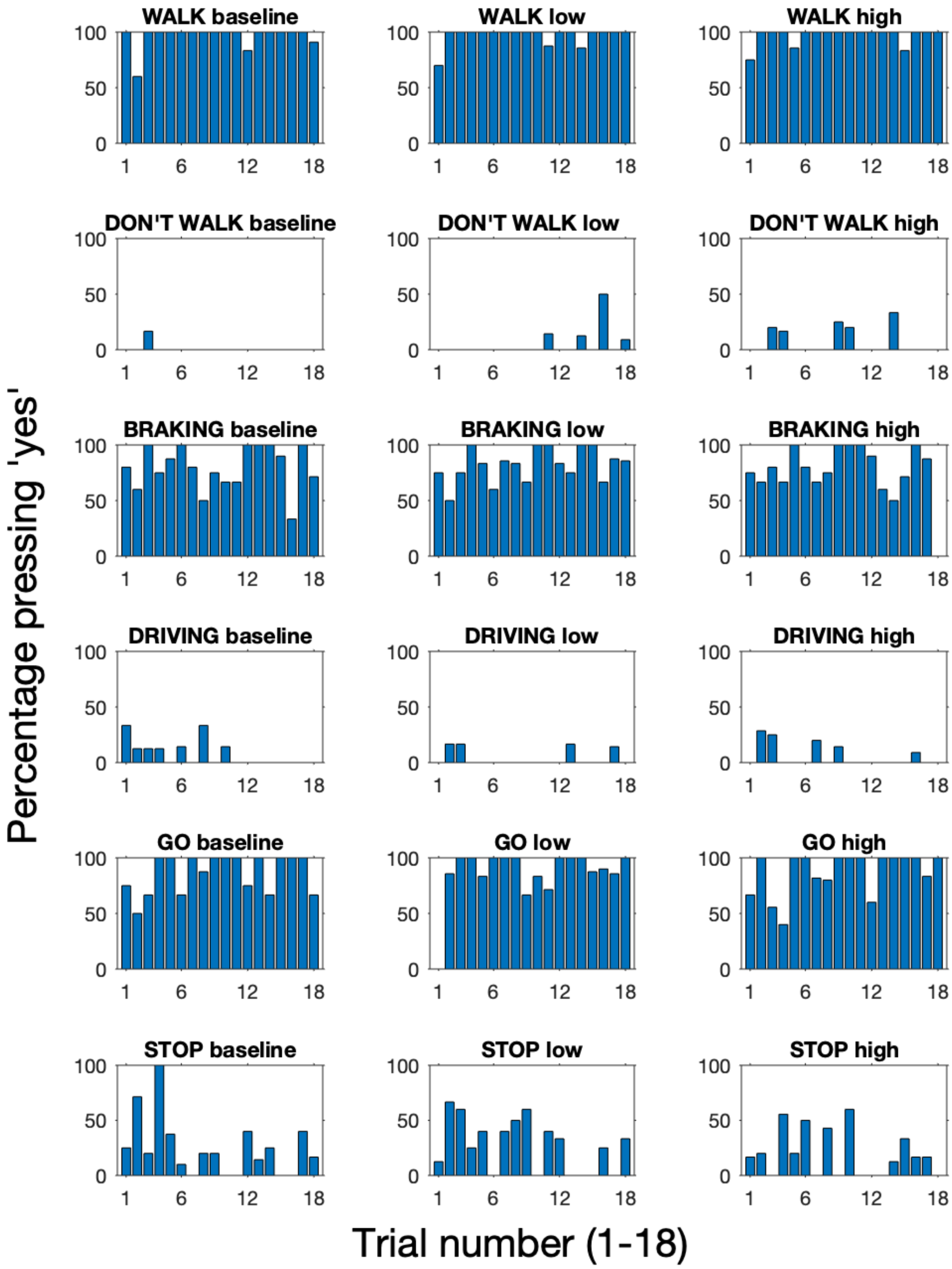


Figure 14. An overview of the percentage pressing 'yes' for each stimuli for each trial during the experiment. The column shows the memory task (baseline, low load and high load) and the rows show the six external Human-Machine Interfaces (eHMIs). For example, the figure shows that the response time for the stimulus 'DON'T WALK' with a high memory load was about one second higher compared to the other trials.

APPENDIX 11 MEAN DIFFERENCE IN RESPONSE TIMES FOR MEMORY LOAD

Table 13. Overview of the difference in response times in miliseconds and percentage for the low and the high load memory task compared to the baseline condition for each eHMI concept. The difference was calculated by subtracting the averaged response times for the baseline condition from the averaged response times for the memory task).

eHMI	*Memory load*	*Memory load*	*Difference (ms)*	*Difference (%)*
WALK	baseline	low	202.7	10.4
WALK	baseline	high	134.4	6.0
DON’T WALK	baseline	low	127.1	3.8
DON’T WALK	baseline	high	46.3	3.8
BRAKING	baseline	low	68.9	0.3
BRAKING	baseline	high	89.0	5.2
DRIVING	baseline	low	236.1	14.2
DRIVING	baseline	high	223.8	16.5
GO	baseline	low	225.5	14.5
GO	baseline	high	263.1	16.2
STOP	baseline	low	-4.1	-3.4
STOP	baseline	high	111.3	6.1

APPENDIX 12 MEAN PUPIL DIAMETER AS A FUNCTION OF TIME FOR MESSAGE PERSPECTIVE AND EHMI

Figure 15 shows the pupil diameter for each eHMI concept averaged over all three memory loads. Altough, as mentioned in the result section, no difference in pupil diameter was found for message perspective, the figure shows that for both yielding and non-yielding the pupil diameter is bigger for the ambiguous messages compared to the egocentric messages at the moment of response. These results suggest that the ambiguity of the message resulted in a higher cognitive load. As can be seen in figure 16 the pupillary responses are identical for the ambiguous messages, indicating that no difference in cognitive load was measured between these message.

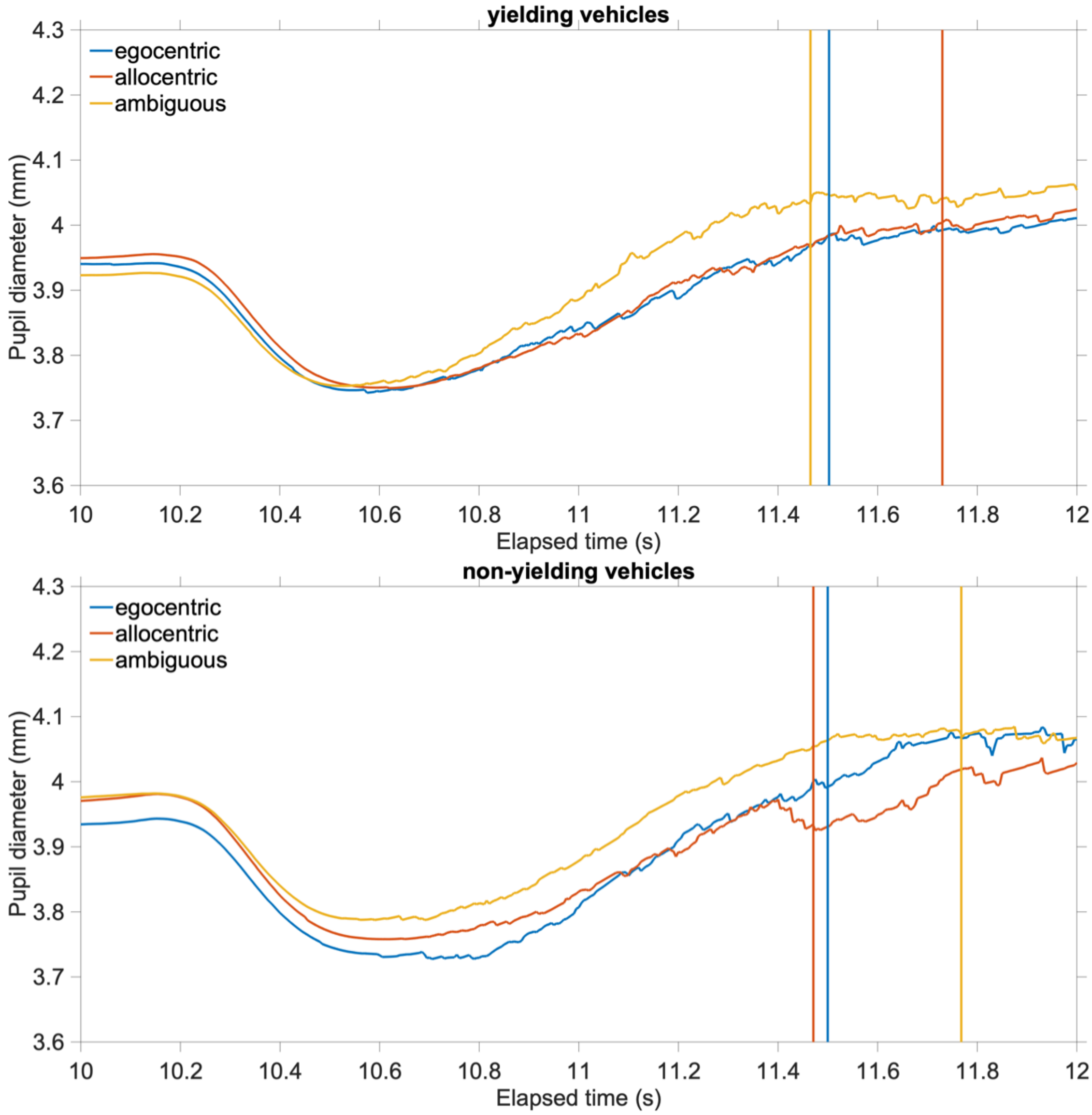


Figure 15. The figures shows the effect of message perspective on the averaged pupil diameter over all three memory loads for yielding and non-yielding vehicles. The colored vertical lines correspond to the mean response time for each message. Egocentric stands for egocentric perspective, allocentric stands for allocentric perspective and ambiguous stands for ambiguous message perspective.

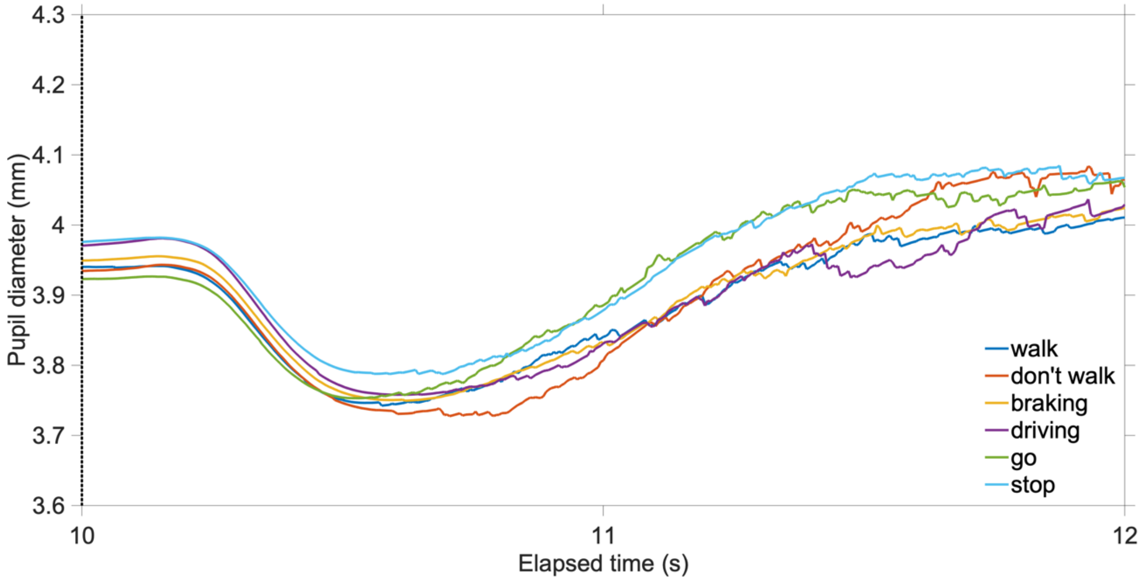


Figure 16. Mean pupil diameter as a function of time for each stimulus. The mean pupil diameter is averaged over all three memory loads.

APPENDIX 13 OVERVIEW OF POST-HOC COMPARISONS OF MEMORY TASK FOR ALL TIME BINS

	MEMORY LOAD	MEMORY LOAD	DIFFERENCE	STDERR	PVALUE	LOWER	UPPER
T_1	baseline	low	-0.001	0.014	1.000	-0.034	0.033
	baseline	high	0.007	0.013	1.000	-0.025	0.040
	low	baseline	0.001	0.014	1.000	-0.033	0.034
T_2	baseline	low	0.002	0.015	1.000	-0.035	0.038
	baseline	high	0.002	0.015	1.000	-0.034	0.037
	low	baseline	-0.002	0.015	1.000	-0.038	0.035
T_3	baseline	low	0.014	0.014	1.000	-0.021	0.048
	baseline	high	-0.026	0.013	0.140	-0.058	0.006
	low	baseline	-0.014	0.014	1.000	-0.048	0.021
T_4	baseline	low	0.011	0.015	1.000	-0.025	0.047
	baseline	high	-0.053	0.012	**< 0.001**	-0.083	-0.022
	low	high	-0.064	0.015	**< 0.001**	-0.100	-0.028
T_5	baseline	low	-0.009	0.013	1.000	-0.040	0.023
	baseline	high	-0.128	0.014	**< 0.001**	-0.163	-0.093
	low	baseline	0.009	0.013	1.000	-0.023	0.040
T_6	baseline	low	-0.039	0.016	0.055	-0.078	0.001
	baseline	high	-0.220	0.019	**< 0.001**	-0.265	-0.174
	low	baseline	0.039	0.016	0.055	-0.001	0.078
T_7	baseline	low	-0.068	0.018	**< 0.001**	-0.111	-0.024
	baseline	high	-0.270	0.022	**< 0.001**	-0.324	-0.216
	low	baseline	0.068	0.018	**< 0.001**	0.024	0.111
T_8	baseline	low	-0.058	0.018	**0.005**	-0.101	-0.015
	baseline	high	-0.205	0.022	**< 0.001**	-0.259	-0.152
	low	baseline	0.058	0.018	**0.005**	0.015	0.101
T_9	baseline	low	-0.010	0.016	1.000	-0.049	0.030
	baseline	high	-0.145	0.018	**< 0.001**	-0.188	-0.101
	low	baseline	0.010	0.016	1.000	-0.030	0.049
T_{10}	baseline	low	-0.020	0.014	0.435	-0.054	0.013
	baseline	high	-0.127	0.015	**< 0.001**	-0.164	-0.091
	low	baseline	0.020	0.014	0.435	-0.013	0.054
T_{11}	baseline	low	-0.019	0.011	0.240	-0.046	0.007
	baseline	high	-0.085	0.013	**< 0.001**	-0.116	-0.054
	low	baseline	0.019	0.011	0.240	-0.007	0.046
T_{12}	baseline	low	-0.036	0.026	0.581	-0.108	0.037
	baseline	high	-0.061	0.028	0.149	-0.138	0.017
	low	baseline	0.036	0.026	0.581	-0.037	0.108

APPENDIX 14: MEAN RESPONSE TIME (MS) FOR PRESSING 'YES' AND 'NO' FOR EACH STIMULUS

Figure 17 shows the mean response times for each stimulus for the pressing 'yes' and 'pressing 'no'. For the ambiguous messages, participants responded faster when interperting the message egocentrically compared allocentrically. These results might point towards an egocentric bias, however, the crossing intentions do suggest but do not provide evidence regarding the perspective taken (Table 14) Therefore, it remains unclear whether longer response times for the allocentric interpretation, are due to an egocentric bias. The averaged response time over all stimuli for pressing 'yes' was 1563 ms (SD = 807) and for pressing 'no' 1655 ms (SD = 803), showing that participants resopnded faster when pressing 'yes' compared to pressing 'no'.

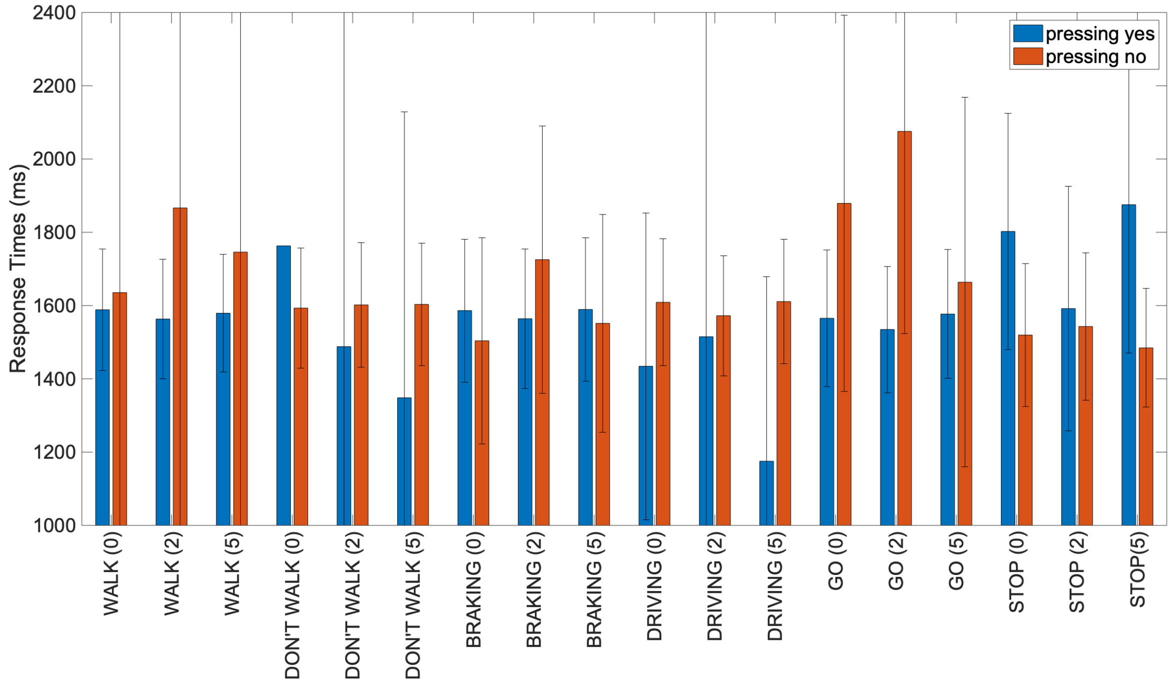


Figure 17. Mean response times for pressing 'yes' and 'no' for each trial. The response times are averaged over the number of participants pressing 'yes' for each stimulus and pressing 'no' for each stimulus. The error bars correspond to the confidence interval.

Table 14. Mean response time for each message pressing 'yes' and 'no'. The mean was calculated by averaging the response times over the number of participants pressing 'yes' and 'no' and the three memory task conditions.

	RT pressing 'yes'		*RT pressing 'no'*	
	Mean	*SD*	*Mean*	*SD*
WALK	1577	792	1758	1053
DON'T WALK	1446	564	1599	815
BRAKING	1580	847	1592	644
DRIVING	1361	560	1597	814
GO	1559	810	1850	818
STOP	1747	786	1514	787

Abbreviations: RT stands for Response Time

APPENDIX 15: INDIVIDUAL RESULTS FOR RESPONSE TIME AND CROSSING INTENTIONS

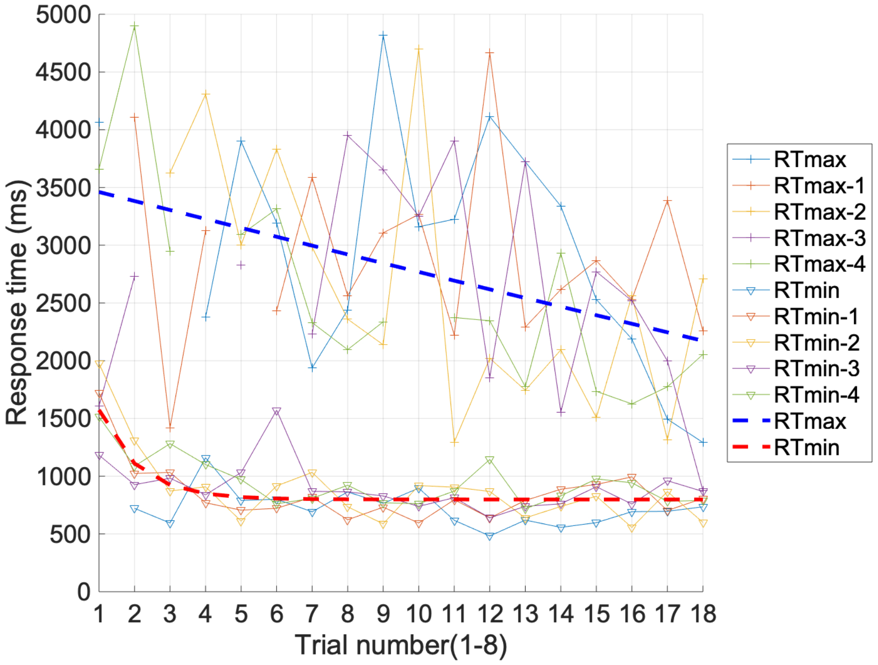


Figure 18. The response times for the the five participants with the fastest response times and the five participants with the slowest response times as a funciton of the trials. The maximum and minimum response times were calculated by averaging the response times over all 18 trials for each participant. The smooth lines represent the learning curves for participants with the longest repsonse times and the participants with the fastest response times. RTmax stands for the participant with the longest reaction time, RTmax-1 for the second longest, RTmax-2 for the third longest, etceteras. The same accounts for RTmin, which stands for the participant with the shortest response times etceteras.

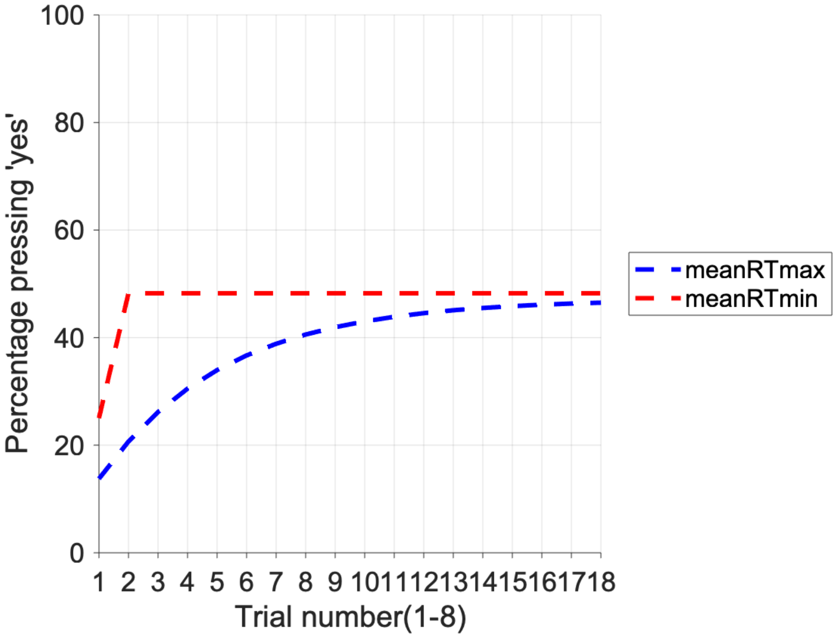


Figure 19. Learning curve for the percentage pressing ‘yes’ for the five participants with the longest response times and the five participants with the shortest response times as a function of the trials.